CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 125, No. 4, Pt. 1 of 2 April 2009

JASA EXPRESS LETTERS

Virtual reflector representation theorem (acoustic medium)	Flavio Poletto, Kees Wapenaar	EL111
Wave refraction at an interface: Snell's law versus Chapman's law	Oleg A. Godin	EL117
***In situ* measurement of permeability**	Lin Lin, M. L. Peterson, Alan R. Greenberg, Benjamin A. McCool	EL123
Plate wave stop-bands in periodically poled lithium niobate	I. V. Ostrovskii, V. A. Klymko, A. B. Nadtochiy	EL129
Predicting echo thresholds from speech onset characteristics	Scott D. Miller, Ruth Y. Litovsky, Keith R. Kluender	EL134
Vocal tract modes based on multiple area function sets from one speaker	Brad H. Story	EL141
A portable infrasound generator	Joseph Park, James Robertson	EL148
Motion of the lips of brass players during extremely loud playing	Samuel Stevenson, Murray Campbell, Seona Bromage, John Chick, Joël Gilbert	EL152
Estimating marine sediment attenuation at low frequency with a vertical line array	Yong-Min Jiang, N. Ross Chapman, Kunde Yang, Yuanliang Ma	EL158
Sand acoustics: The effective density fluid model, Pierce/Carey expressions, and inferences for porous media modeling	Kevin L. Williams	EL164
Statistical estimation of source location in presence of geoacoustic inversion uncertainty	Chen-Fen Huang, Peter Gerstoft, William S. Hodgkiss	EL171

LETTERS TO THE EDITOR

Monitoring stress related velocity variation in concrete with a 2 $\times 10^{-5}$ relative resolution using diffuse ultrasound (L)	Eric Larose, Stephen Hall	1853
Comment on "Measurement of the frequency dependence of the ultrasonic parametric threshold amplitude for a fluid-filled cavity" (L)	M. Strasberg	1857
Time domain formulation of the equivalent fluid model for rigid porous media (L)	Olga Umnova, Diego Turo	1860
Prediction of the acoustic field due to a point source over a porous layer (L)	Jean F. Allard	1864

GENERAL LINEAR ACOUSTICS [20]

A filtered convolution method for the computation of acoustic wave fields in very large spatiotemporal domains	Martin D. Verweij, Jacob Huijssen	1868

(Continued)

CONTENTS—*Continued from preceding page*

NONLINEAR ACOUSTICS [25]

A micro-machined source transducer for a parametric array in air	Haksue Lee, Daesil Kang, Wonkyu Moon	1879
On the Larsen effect to monitor small fast changes in materials	Oleg I. Lobkis, Richard L. Weaver	1894
Wideband nonlinear time reversal seismo-acoustic method for landmine detection	Alexander Sutin, Brad Libbey, Laurent Fillinger, Armen Sarvazyan	1906

AEROACOUSTICS, ATMOSPHERIC SOUND [28]

Finite difference time domain electroacoustic model for synthetic jet actuators including nonlinear flow resistance	Gerben Kooijman, Okke Ouweltjes	1911

UNDERWATER SOUND [30]

The interference component of the acoustic field corresponding to the Long-Range Ocean Acoustic Propagation Experiment	Natalie S. Grigorieva, Gregory M. Fridman, James A. Mercer, Rex K. Andrew, Michael A. Wolfson, Bruce M. Howe, John A. Colosi	1919
Effects of incident field refraction on scattered field from vertically extended cylindrical targets in range-dependent ocean waveguides	Elizabeth T. Küsel, Purnima Ratilal	1930
A simplified approach to solving finite rough surface scattering problems	J. A. Fawcett	1937
Finite-difference modeling of the monopole acoustic logs in a horizontally stratified porous formation	Wei Guan, Hengshan Hu, Xiao He	1942
Laboratory investigation of the acoustic response of seagrass tissue in the frequency band 0.5–2.5 kHz	Preston S. Wilson, Kenneth H. Dunton	1951
Retrieval of Green's functions of elastic waves from thermal fluctuations of fluid-solid systems	Oleg A. Godin	1960
Revisiting experimental methods for studies of acidity-dependent ocean sound absorption	Timothy F. Duda	1971
Estimating cetacean population density using fixed passive acoustic sensors: An example with Blainville's beaked whales	Tiago A. Marques, Len Thomas, Jessica Ward, Nancy DiMarzio, Peter L. Tyack	1982

TRANSDUCTION [38]

Fundamental constraints on the performance of broadband ultrasonic matching structures and absorbers	O. Acher, J. M. L. Bernard, P. Maréchal, A. Bardaine, F. Levassort	1995
The rotary subwoofer: A controllable infrasound source	Joseph Park, Milton Garcés, Bruce Thigpen	2006
A low-noise differential microphone inspired by the ears of the parasitoid fly *Ormia ochracea*	R. N. Miles, Q. Su, W. Cui, M. Shetye, F. L. Degertekin, B. Bicen, C. Garcia, S. Jones, N. Hall	2013

STRUCTURAL ACOUSTICS AND VIBRATION [40]

Determination of the random anisotropic elasticity layer using transient wave propagation in a fluid-solid multilayer: Model and experiments	Christophe Desceliers, Christian Soize, Quentin Grimal, Maryline Talmant, Salah Naili	2027
Weakly nonlinear oscillations of a compliant object buried in soil	Evgenia A. Zabolotskaya, Yurii A. Ilinskii, Mark F. Hamilton	2035
Elastodynamic wave scattering by finite-sized resonant scatterers at the surface of a horizontally layered halfspace	Geert Lombaert, Didier Clouteau	2041

(Continued)

CONTENTS—*Continued from preceding page*

A wave-based finite element analysis for acoustic transmission in fluid-filled elastic waveguides	Andrew T. Peplow	2053

ARCHITECTURAL ACOUSTICS [55]

A correction of random incidence absorption coefficients for the angular distribution of acoustic energy under measurement conditions	Cheol-Ho Jeong	2064
Increase in voice level and speaker comfort in lecture rooms	Jonas Brunskog, Anders Christian Gade, Gaspar Payá Bellester, Lilian Reig Calbo	2072
A general wave decomposition formula for the measurement of normal incidence sound transmission loss in impedance tube	Yacoubou Salissou, Raymond Panneton	2083

ACOUSTIC SIGNAL PROCESSING [60]

A realization of sound focused personal audio system using acoustic contrast control	Ji-Ho Chang, Chan-Hui Lee, Jin-Young Park, Yang-Hann Kim	2091
Beamforming with microphone arrays for directional sources	Christian Bouchard, David I. Havelock, Martin Bouchard	2098
Basic theory and properties of statistically optimized near-field acoustical holography	Jørgen Hald	2105
Multiple-point statistical room correction for audio reproduction: Minimum mean squared error correction filtering	Fredrik Lingvall, Lars-Johan Brännmark	2121

PHYSIOLOGICAL ACOUSTICS [64]

Estimating the operating point of the cochlear transducer using low-frequency biased distortion products	Daniel J. Brown, Jared J. Hartsock, Ruth M. Gill, Hillary E. Fitzgerald, Alec N. Salt	2129
Inverse-solution method for a class of non-classical cochlear models	Egbert de Boer, Alfred L. Nuttall	2146

PSYCHOLOGICAL ACOUSTICS [66]

Judging the similarity of soundscapes does not require categorization: Evidence from spliced stimuli	Jean-Julien Aucouturier, Boris Defreville	2155
The masking of interaural delays	Andrew J. Kolarik, John F. Culling	2162
Precursor effects on behavioral estimates of frequency selectivity and gain in forward masking	Skyler G. Jennings, Elizabeth A. Strickland, Michael G. Heinz	2172
Auditory stream formation affects comodulation masking release retroactively	Torsten Dau, Stephan Ewert, Andrew J. Oxenham	2182
Can temporal fine structure represent the fundamental frequency of unresolved harmonics?	Andrew J. Oxenham, Christophe Micheyl, Michael V. Keebler	2189
Release from informational masking in children: Effect of multiple signal bursts	Lori J. Leibold, Angela Yarnell Bonino	2200
Maximal azimuthal resolution needed in measurements of head-related transfer functions	Xiao-Li Zhong, Bo-Sun Xie	2209
Specificity of binaural perceptual learning for amplitude modulated tones: A comparison of two training methods	Daniel Kumpik, Jeremy Ting, Robert A. A. Campbell, Jan W. H. Schnupp, Andrew J. King	2221
A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function	Alan Kan, Craig Jin, André van Schaik	2233

(Continued)

CONTENTS—*Continued from preceding page*

Wind noise in hearing aids with directional and omnidirectional microphones: Polar characteristics of behind-the-ear hearing aids	King Chung, Luc Mongeau, Nicholas McKibben	2243
Improved fundamental frequency coding in cochlear implant signal processing	Matthias Milczynski, Jan Wouters, Astrid van Wieringen	2260

SPEECH PRODUCTION [70]

Characteristics of air puffs produced in English "pa": Experiments and simulations	Donald Derrick, Peter Anderson, Bryan Gick, Sheldon Green	2272
Measurement of liquid and solid component parameters in canine vocal fold lamina propria	Robert Phillips, Yu Zhang, Megan Keuler, Chao Tao, Jack J. Jiang	2282
An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan	Daniel Recasens, Aina Espinosa	2288
Attenuation of vocal responses to pitch perturbations during Mandarin speech	Hanjun Liu, Yi Xu, Charles R. Larson	2299
Place assimilation across words in running speech: Corpus analysis and perception	Frank Zimmerer, Henning Reetz, Aditi Lahiri	2307

SPEECH PERCEPTION [71]

Estimating vowel formant discrimination thresholds using a single-interval classification task	Eric Oglesbee, Diane Kewley-Port	2323
Speech intelligibility in background noise with ideal binary time-frequency masking	DeLiang Wang, Ulrik Kjems, Michael S. Pedersen, Jesper B. Boldt, Thomas Lunner	2336
Perception of allophonic cues to English word boundaries by Japanese second language learners of English	Kikuyo Ito, Winifred Strange	2348
Processing of fast speech by elderly listeners	Esther Janse	2361

SPEECH PROCESSING AND COMMUNICATION SYSTEMS [72]

A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data	Richard E. Turner, Thomas C. Walters, Jessica J. M. Monaghan, Roy D. Patterson	2374
Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs	Geoffrey Stewart Morrison	2387

MUSIC AND MUSICAL INSTRUMENTS [75]

Absolute pitch among students in an American music conservatory: Association with tone language fluency	Diana Deutsch, Kevin Dooley, Trevor Henthorn, Brian Head	2398
Trombone bore optimization based on input impedance targets	Alistair C. P. Braden, Michael J. Newton, D. Murray Campbell	2404

BIOACOUSTICS [80]

A bulk modulus dependent linear model for acoustical imaging	Jean Martial Mari, Thierry Blu, Olivier Bou Matar, Michael Unser, Christian Cachard	2413
Magnetic resonance imaging of boiling induced by high intensity focused ultrasound	Tatiana D. Khokhlova, Michael S. Canney, Donghoon Lee, Kenneth I. Marro, Lawrence A. Crum, Vera A. Khokhlova, Michael R. Bailey	2420

(Continued)

CONTENTS—*Continued from preceding page*

Forward-masking based gain control in odontocete biosonar: An evoked-potential study — Alexander Ya. Supin, Paul E. Nachtigall, Marlee Breese — 2432

Modeling cumulative sound exposure around marine seismic surveys — Christine Erbe, Andrew R. King — 2443

ERRATA

Erratum: "Born iterative reconstruction using perturbed-phase field estimates " [J. Acoust. Soc. Am., 124, 2353–2363 (2008)] — Jeffrey P. Astheimer, Robert C. Waag — 2452

ACOUSTICAL NEWS—USA — 2453

USA Meeting Calendar — 2460

ACOUSTICAL NEWS—INTERNATIONAL — 2461

International Meeting Calendar — 2461

REVIEWS OF ACOUSTICAL PATENTS — 2462

CUMULATIVE AUTHOR INDEX — 2479

Virtual reflector representation theorem (acoustic medium)

Flavio Poletto
Instituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS), Borgo Grotta Gigante 42/c, 34010 Sgonico (Trieste), Italy
fpoletto@ogs.trieste.it

Kees Wapenaar
Department of Geotechnology, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands
c.p.a.wapenaar@tudelft.nl

Abstract: The virtual reflector method simulates new seismic signals by processing traces recorded by a plurality of sources and receivers. The approach is based on the crossconvolution of the recorded signals and makes it possible to obtain the Green's function of virtual reflected signals as if in the position of the receivers (or sources) there were a reflector, even if said reflector is not present. This letter presents the virtual reflector theory based on the Kirchhoff integral representation theorem for wave propagation in an acoustic medium with and without boundary and a generalization to variable reflection coefficients for scattered wavefields.

PACS numbers: 43.20.Bi, 43.20.El, 43.60.Tj [AE]
Date Received: November 19, 2008 **Date Accepted:** January 5, 2009

1. Introduction

In recent years innovative techniques have been introduced in seismics and acoustics to simulate new signals by processing the measurements obtained by a plurality of sources and receivers. These methods are generally classified as "Green's function retrieval" or "seismic interferometry" (SI) and are essentially based on crosscorrelation of traces at different receivers and summation of the crosscorrelated signals over the space of the sources generating the wavefields. In this process, it is not necessary to know the sources and the medium. From a physical point of view, crosscorrelation (or convolution by time reversal) is equivalent to removing the common propagation effects.[1,2] A great advantage of this approach is that also signals from unknown random-phase sources are shaped into interpretable signals because of phase subtraction. In some applications, deconvolution is used instead of correlation to compensate also for source amplitude.[3] Under proper conditions, SI allows one to recover the signal ideally produced by a virtual source located at the position of a receiver.[4] This technique has important applications for ultrasonics,[5] underwater acoustics,[6] seismic exploration, and passive seismics. The method makes it possible to recover the local receiver-to-receiver Earth's response, i.e., the Green's function, and provides virtual sources where it is not possible or difficult to use a real seismic source, such as in borehole and sea-floor marine surveys. Interferometry has been the subject of several theoretical[1–3,5–7] and application studies, and nowadays it is used for seismic exploration purposes.[8,9] With an approach similar to interferometry, Poletto[10] proposed the virtual reflector (VR) method,[11] which also simulates new seismic signals by processing recorded traces from a plurality of sources and receivers, with the following differences with respect to interferometry. The method is based on crossconvolution (different from convolution with time reversal) of source (receiver) signals and subsequent summation in the domain of the receivers (sources). The VR method substantially performs the composition of the filtering effects in the wave propagation. To be effective, the method needs to deal with source signals with transient wavelets. The novel approach makes it possible to obtain virtual reflected signals as if at the

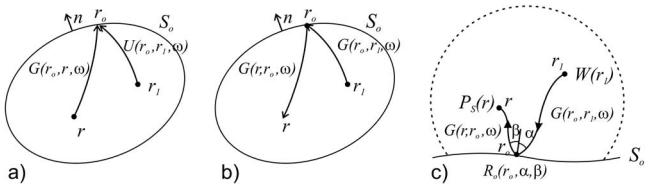


Fig. 1. (a) Acoustic model for VR representation. Two sources are located at r and r_1 encompassed by the surface S_o. (b) Composition of propagation effects and surface integration to obtain the VR signal. (c) The scattered field from reflector S_o is represented by a boundary integral over the reflector. The reflector coefficient $R_o(r_o,\alpha,\beta)$ depends on the reflector geometry and acquisition configuration.

position of receivers (or sources) there were a real reflector even if said reflector is not present there. In this letter we re-formulate the VR theory for an acoustic inhomogeneous medium by using the Kirchhoff–Helmholtz (KH) integral representation theorem.

2. Theory

The scalar fields in an inhomogeneous medium can be expressed by using the Green's integral theorem for scalar functions U and G. Let S be the total surface encompassing a volume of interest V, and n the outward normal direction to the surface. The Green's integral theorem states that

$$\int_V dV[G\nabla^2 U - U\nabla^2 G] = \int_S dS[G\,\nabla\, U - U\,\nabla\, G]\cdot n. \tag{1}$$

The functions U and G are assumed to be the Fourier transforms of causal functions, and in the common use they represent the propagating wavefield and the medium response to an impulse, Green's function, respectively. Let S_o be the outer surface enclosing the volume of interest V. The VR signal is the result of the composition of scalar fields from different sources included in the volume V surrounded by receivers located at outer surface S_o (Ref. 11). Assume that two point sources at internal points r_1 and at r inside S_o [Fig. 1(a)] generate the scalar fields U and G, respectively. They can be expressed as[12–14]

$$U(r,r_1,\omega) - G(r_1,r,\omega) = \frac{1}{4\pi}\int_{S_o} dS_o\left[G(r_o,r,\omega)\frac{\partial U(r_o,r_1,\omega)}{\partial n} - U(r_o,r_1,\omega)\frac{\partial G(r_o,r,\omega)}{\partial n}\right], \tag{2}$$

where $U(r_o,r_1,\omega)$ is the scalar field generated at r_1 and recorded along the surface S_o, r_o is location along the surface, $G(r_o,r,\omega)$ is the Green's function from r to r_o, $\partial/\partial n$ is the normal differentiation operator acting on r_o at surface S_o, and ω is the angular frequency. The advantage of the integral representation is that a suitable propagation Green's function G can be chosen to synthesize a particular U wavefield (and vice versa). Equation (2), with fields U and G satisfying different boundary conditions (BCs), was exploited in developing the reverse-time approach to reflection seismic imaging (e.g., Refs. 15 and 16). When U and G are subject to the same BCs on S_o, the integral vanishes. In this case from Eq. (2), we obtain $G=U$. Conversely, under proper BCs, different for U and G, the integral is in general different from zero. This result corresponds to the fact that, even when the wavefield represented by G from r to r_1 (from r_1 to r if we use reciprocity) is subtracted from U in Eq. (2), part of the synthesized wavefield still remains. If we assume that the medium outside S_o is homogeneous and there are no reflecting objects outside S_o (they may be on S_o), Eq. (2) can be used to represent a perfect VR at S_o. We define a new (VR) wavefield as $V=U-G$. This result comes from the assumption of reflecting BCs on the close surface S_o for G but not for U (or vice versa). In V we subtract direct arrivals and reflections. This reasoning is in agreement with the fact that the VR method provides *only*, however not necessarily *all*, reflected waves from S_o (Ref. 11). The scalar function V,

when derived by Eq. (2), can be used to represent a perfect VR of unit (or constant) reflection coefficient. A generalization can be obtained by including a variable reflection coefficient function, according to a boundary integral representation for the response of a reflector.

Under the assumption that the scalar field $U(r_o,r_1,\omega)$ is produced from a unit source at point r_1 included in the space encompassed by S_o [Fig. 1(a)], said propagation function represents, by definition, the medium response. In this case we can express also the scalar quantity recorded on S_o as $U(r_o,r_1,\omega)=G(r_o,r_1,\omega)$, where $G(r_o,r_1,\omega)$ is the Green's function from r_1 to r_o, and substitute G for U in Eq. (2), by using V, which gives

$$V(r,r_1,\omega)=\frac{1}{4\pi}\int_{S_o} dS_o\left[G(r_o,r,\omega)\frac{\partial G(r_o,r_1,\omega)}{\partial n}-G(r_o,r_1,\omega)\frac{\partial G(r_o,r,\omega)}{\partial n}\right]. \quad (3)$$

In this representation it is intended that the reflecting BC on S_o is applied to only one of the two terms of the sum in the integrand at the right hand side.

3. Synthesis of virtual reflector

Equation (3) allows us to synthesize a VR wavefield with constant reflection coefficient under different BCs. We first analyze the case in which the model contains an equivalent medium with a reflecting interface at the boundary. We consider two cases in which $G(r_o,r,\omega)$ is chosen with a perfect reflecting BC on S_o. The first reflecting BC on S_o for $G(r_o,r,\omega)$ calculated for a given r is $\partial G(r_o,r,\omega)/\partial n=0$ (*Neumann* BC). In this case the contrast medium is perfectly rigid and reflecting (reflection coefficient $R=+1$). Equation (3) becomes

$$V_R(r,r_1,\omega)=\frac{-\imath\omega}{4\pi c}\int_{S_o} dS_o G(r_o,r,\omega)G(r_o,r_1,\omega)\cos(\alpha_o), \quad (4)$$

where we have used the approximation $\partial/\partial n\cong-\imath\omega\cos(\alpha_o)/c$ for the normal derivative, $\imath=\sqrt{-1}$, c is the medium velocity, and α_o is the angle between the ray of $G(r_o,r_1,\omega)$ and the normal at r_o on S_o. The second reflecting BC for $G(r_o,r,\omega)$ on S_o is $G(r_o,r,\omega)=0$ (*Dirichlet* BC). In this case, the contrast medium is vacuum, corresponding to a perfectly-reflecting free surface ($R=-1$). Equation (3) becomes (see Ref. 15)

$$V_F(r,r_1,\omega)=\frac{-1}{4\pi}\int_{S_o} dS_o G(r_o,r_1,\omega)\frac{\partial G(r_o,r,\omega)}{\partial n}. \quad (5)$$

Deriving an equation similar to Eq. (4) from Eq. (5) is strictly not valid for the free surface condition because $G(r_o,r,\omega)$ vanishes on S_o. Conversely, if we substitute the wavefield of the corresponding model without boundary at S_o for the wavefield of the model with boundary, to avoid vanishing of G, and use the same normal-derivative approximation, we obtain as an approximation a virtual relation similar to Eq. (4) with opposite sign.

From Eq. (4) we obtain the VR signal for the acoustic medium *with* reflecting boundary. According to the concept of the VR method,[10] the integral in Eq. (4) is the integral on S_o of the crossconvolution[1] of the non-zero scalar quantities (a factor of 4 is here neglected) measured by receivers located on S_o and produced by unit sources located at r and r_1, respectively [Fig. 1(a)]. Because we consider measured wavefields in an unknown medium configuration, we ignore the factor $\cos(\alpha_o)$. Moreover, to interpret Eq. (4) in terms of propagating wavefields, we use reciprocity and exchange source (ideal image of a reflected source) and receiver at r and r_o (or equivalently at r_1 and r_o). Using the reciprocity condition $G(r,r_o,\omega)=G(r_o,r,\omega)$, the VR representation equation for the equivalent medium with reflecting boundary becomes

$$V_{RF}(r,r_1,\omega) = \frac{\mp \imath\omega}{4\pi c}\int_{S_o} dS_o G(r,r_o,\omega)G(r_o,r_1,\omega), \tag{6}$$

which means that the VR signal is obtained by composing the propagated signals from a unit source at point r_1 to receiver in r_o and from r_o to r (or vice versa) and by integrating over the receiver surface the convolution results [Fig. 1(b)]. Consider now the case in which a reflector is *not* present at S_o boundary (transparent surface). The function $\hat{G}$ *without* reflecting boundary at S_o is related to the function G of the model with boundary by $G(r_o,r_j,\omega)=\hat{G}(r_o,r_j,\omega)+\Delta G(r_o,r_j,\omega)$, where ΔG represents the additional effects due to the presence of the reflecting boundary. In other words, $\hat{G}$ is an approximation of the Green's function G with reflecting boundary, in the sense that $\hat{G}$ does not contain reflections from S_o. The virtual reflections $\hat{V}(r,r_1,\omega)$ from S_o without reflecting boundary are synthesized by substituting the signal $\hat{G}$ for G in Eq. (6). Result (4) [or (6)] and its unbounded approximations are equivalent to the summation of crossconvolutions of the measured signals over the receiver (or source) space as proposed by Ref. 10. For selected events it holds $\hat{V}(r,r_1,\omega)\propto V(r,r_1,\omega)$ since the VR method synthesizes only but not all the reflections from S_o (Ref. 11), without the effects of the additional term ΔG.

The derivation of a generalization of Eq. (6) for the scattered wavefield is based on work of Refs. 17, 12, and 18. For the configuration in Fig. 1(c), the scattered wavefield at an observation point r above the reflector is given by

$$P_S(r,r_1,\omega) = \int_{S_o} dS_o G(r,r_o,\omega)R_o(r_o,\alpha,\beta)G(r_o,r_1,\omega)W(r_1,\omega), \tag{7}$$

where $R_o(r_o,\alpha,\beta)$ is the angle-dependent reflection operator and $W(r_1,\omega)$ is the source signal injected at r_1. In the most general case the reflection operator is a pseudodifferential operator (Ref. 17), but for the VR method we approximate it by a high-frequency angle-dependent reflection coefficient (Refs. 12 and 18). Here $P_S(r,r_1,\omega)$ is the scattered wavefield of the true medium, including the reflector, whereas $G(r,r_o,\omega)$ and $G(r_o,r_1,\omega)$ are Green's functions in one-and-the-same reference medium, which is identical to the true medium above the reflector but continues without a jump in the medium parameters below the reflector. The step to the VR principle is again easily made: assume that in the true medium, S_o is not a reflector but a surface with receivers (or sources), whereas r and r_1 both denote sources (or receivers), then the Green's functions with interchanged coordinates in one of them are obtained by measurements as proposed in the previous sections. Substituting these measurements in Eq. (7), inserting a user defined reflection coefficient, and evaluating the integral gives the VR response P_S. In its simplest form $R_o(r_o,\alpha,\beta)$ can be taken equal to $2i\omega R(r_o)/[c(r_o)\rho(r_o)]$, where ρ is the density of the medium. In particular, by using $R=\mp 1$ with normal incidence approximation and unit source signal $W(r_1,\omega)=1$, from Eq. (7) we obtain again Eq. (6), apart from a scaling factor.

All the reasoning made for the VR representation by sources surrounded by receivers holds, using reciprocity, also for the representation with receivers surrounded by sources.

4. Examples, summary, and conclusions

Figure 2 shows the model used for a two dimensional acoustic simulation by a finite-differences code, where a background homogeneous medium with compressional velocity 2 km/s includes a diffraction body at point D. Synthetic signals are calculated with and without the presence of a reflecting boundary at the circumference (surface S_o) of a circle of radius of 1.8 km. The outer hard-contrast medium velocity is 20 km/s. The signals of two sources, S_1 and S_2, are recorded by 360 receivers located on S_o. Figure 2(c) compares the simulated signals from source S_1 to a control receiver at the position of the source S_2 in the model (1) without and (2) with circular

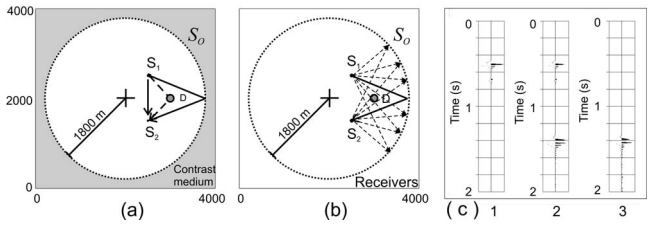


Fig. 2. Synthetic model: (a) With contrast medium at circular boundary S_o. The arrows indicate the direct boundary reflection and diffraction (dashed line) arrivals from the source S_1 to a control receiver at S_2. (b) Representation of the VR arrival (solid line) from the source S_1 to the source S_2. The VR signal is reflected by a VR at S_o and is obtained by processing the signals (dashed arrows) recorded by the receivers on the circle without the contrast medium. (c) Synthetic traces. (1) Control signal without reflector. (2) Control signal with reflector. (3) VR signal from S_1 to S_2.

reflecting boundary and (3) the VR signal calculated by performing the crossconvolution of the signals from the two sources and by integrating the crossconvolutions over the receiver space S_o in the model without reflection boundary.

To summarize, we calculated the KH integral representation of the virtual signal produced by a reflecting surface surrounding two points where sources are located. We showed that the reflection representation is equivalent to composing the wavefields of the two sources recorded on the surface. The analysis demonstrated that halfside of the KH integral represents the crossconvolution term of the VR signal. We generalized the representation for the scattered wavefield from a VR with variable reflection coefficients. Note that the VR method is complementary to SI. The main feature of SI is that the response of a virtual source can be generated without knowing the medium; all that is required is a receiver illuminated from many directions, at the location where one wants to create the virtual source. The VR method enables the generation of the reflection response of a VR without knowing the medium, as long as there are receivers (or sources) distributed along the surface at which one wants to create the VR.

The VR signal synthesis with an ideal reflector has several applications with exploration data.[11] Reference 19 shows examples of identification of and removal of surface reflection events in synthetic marine data and performs the analysis of virtual signals with real borehole data. Other potential applications are signal phase analysis and recovery of the source delay. In these applications the VR signal is used in combination with the interferometry one, provided that the VR and the SI results be obtained by using data with similar source-receiver geometry. For these methods a sufficient spatial sampling of receivers (sources) is required. From a practical point of view, this condition can be reasonably achieved in exploration seismics.

References and links

[1]The product $G(r_o,r,\omega)G(r_o,r_1,\omega)$ is the Fourier transform of a temporal crossconvolution. In this paper we loosely refer to such products as (cross-)convolutions.

[1]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).
[2]K. Wapenaar, "Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation," Phys. Rev. Lett. **93**, 254301 (2004).
[3]R. Snieder, J. Sheiman, and R. Calvert, "Equivalence of the virtual-source method and wave-field deconvolution in seismic interferometry," Phys. Rev. E **73**, 066620 (2006).
[4]R. W. Calvert, "Seismic imaging a subsurface formation," U.S. Patent No. 6,747,915 (2004).
[5]E. Larose, L. Margerin, A. Derode, B. van Tiggelen, M. Campillo, N. Shapiro, A. Paul, L. Stehly, and M. Tanter, "Correlation of random wave fields: An interdisciplinary review," Geophysics **71**, SI11–SI21 (2006).
[6]P. Roux, W. A. Kuperman, and the NPAL Group, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[7]K. Wapenaar and J. Fokkema, "Green's function representations for seismic interferometry," Geophysics **71**, SI33–SI46 (2006).

[8]A. Bakulin, A. Mateeva, K. Mehta, P. Jorgensen, J. Ferrandis, I. Sinha Herhold, and J. Lopez, "Virtual source applications to imaging and reservoir monitoring," The Leading Edge **26**, 732–740 (2007).
[9]B. E. Hornby and J. Yu, "Interferometric imaging of a salt flank using walkaway VSP data," The Leading Edge **26**, 760–763 (2007).
[10]F. Poletto, "Method of detection and/or processing of seismic signals," Patent (pending).
[11]F. Poletto and B. Farina, "Synthesis of a virtual reflector by processing recorded seismic signals," Extended Abstracts, 70th EAGE Conference, Rome (2008).
[12]N. Bleistein, *Mathematical Methods for Wave Phenomena* (Academic, New York, 1984).
[13]A. T. De Hoop, *Handbook of Radiation and Scattering Waves: Acoustic Waves in Fluids, Elastic Waves in Solids, Electromagnetic Waves* (Academic, London, 1995).
[14]M. V. De Hoop and A. T. De Hoop, "Wave-field reciprocity and optimization in remote sensing," Proc. R. Soc. London, Ser. A **456**, 641–682 (2000).
[15]W. A. Schneider, "Integral formulation for migration in two and three dimensions," Geophysics **43**, 49–76 (1978).
[16]C. Esmersoy and M. Oristaglio, "Reverse-time wave-field extrapolation, imaging, and inversion," Geophysics **53**, 920–931 (1988).
[17]M. Taylor, "Reflection of singularities of solutions to systems of differential equations," Commun. Pure Appl. Math. **28**, 457–478 (1975).
[18]C. P. A. Wapenaar and A. J. Berkhout, "Representation of seismic reflection data. Part I: State of affairs," J. Seism. Explor. **2**, 123–131 (1993).
[19]F. Poletto and B. Farina, "Seismic virtual reflector—Synthesis and composition of virtual wavefields." Expanded Abstracts, 78th Annual International Meeting, SEG, Las Vegas, Nevada, 9–14 November 2008, Paper PSC 1.7, pp. 1367–1371.

Wave refraction at an interface: Snell's law versus Chapman's law

Oleg A. Godin
CIRES, University of Colorado, 216 UCB, Boulder, Colorado 80309-0216
and NOAA/Earth System Research Laboratory, 325 Broadway, Boulder, Colorado 80305
oleg.godin@noaa.gov

Abstract: Energy streamlines provide insights into mechanisms of wave propagation and scattering and are often utilized to visualize wave fields. In contrast to rays, which are essentially an asymptotic, short-wave concept, energy streamlines adequately represent arbitrary wave fields. However, the usefulness of energy streamlines in studies of wave fields is limited by the fact that, unlike rays, no general laws governing energy streamline refraction are known. Here, a simple refraction law is derived for energy streamlines of acoustic and linearly polarized electromagnetic waves. It is shown that analysis of energy streamlines provides a helpful supplementary perspective on wave transmission through interfaces.

PACS numbers: 43.20.El, 43.20.Bi, 41.20.Jb, 43.58.Fm [AN]
Date Received: December 22, 2008 **Date Accepted:** January 15, 2009

1. Introduction

The energy streamline is defined as a line tangent to which at every point is parallel to the vector of power-flux density at that point. This definition is analogous to the definition of streamline in fluid mechanics as the curve everywhere tangent to the instantaneous direction of the local fluid velocity. Energy streamlines describe energy flow much like conventional streamlines describe mass flow in fluid mechanics.[1] The concept of wave energy streamlines and streamline tubes was introduced by Umov in 1873 and employed in his studies of energy fluxes in fluids and solids (Ref. 2, pp. 201–226 and 535–539). He made use of energy conservation to relate wave intensity to variations in the cross-sectional area of the streamline tube (Ref. 2, pp. 203–204). Umov referred to energy streamlines as "dynamic lines." The modern term was coined by Waterhouse *et al.*[3] Energy streamlines can be defined for instantaneous power flux and for power flux averaged over a wave cycle. For continuous waves in a stationary medium, streamlines of cycle-averaged power flux are independent of time and therefore are the paths along which cycle-averaged wave energy is transmitted from a wave source. A detailed comparison of transmission paths for instantaneous and time-averaged wave energies is presented in Ref. 4.

Energy streamlines reveal intricate details of energy flow in wave fields. A surge in interest to acoustic energy streamlines began in the 1980s (Refs. 3 and 5–7) with the wide availability and application of sound-intensity meters, which measure the vector of power-flux density.[7] As discussed by Fahy (Ref. 7, p. 3), the vector nature of acoustic intensity measurements, as opposed to measurements of scalar acoustic pressure, revolutionized the practice of noise control. Modern applications of wave energy streamlines include studies of wave front dislocations,[8,9] source localization,[7,10] energy vortices in compressible fluids[3,6] and elastic waveguides,[11,12] nanoscale thermal radiation,[13,14] photon tunneling,[13] and bounded beam diffraction, including the Goos–Hänchen shift.[13] Examples of energy streamlines for various wave fields can be found in Refs. 5, 10–14, and 19–22.

Wave energy streamlines are often viewed as an enhancement of rays or as "exact rays"[15] that extend intuitive, graphic portrayal of the wave field inherent to ray theory to evanescent waves and diffracted wave fields, for which the ray theory does not offer an adequate representation. Application of energy streamlines to analyses of specific problems has been impeded by the fact that, unlike rays, no simple laws that govern refraction of streamlines in

volume and at interfaces were known. Abrupt change in energy streamline direction at an interface was observed[10] in results of numerical simulations of acoustic energy streamlines in the problem of spherical wave refraction at a fluid-fluid interface. A remarkably simple refraction law for the acoustic energy streamlines was derived by Chapman[10] from consideration of plane wave transmission through a plane interface of homogeneous fluids. While refraction of rays at an interface, according to Snell's law, is controlled by refraction index, refraction of acoustic energy streamlines, according to Chapman's law, is controlled by the ratio of mass densities of the two fluids in contact.

In this paper, we show that Chapman's law of acoustic energy streamline refraction remains valid at an arbitrary interface in a three-dimensionally inhomogeneous fluid. We also derive similar refraction laws for energy streamlines of linearly polarized electromagnetic waves. The latter refraction laws are valid at interfaces in layered media and in other problems where electromagnetic waves with two orthogonal linear polarizations propagate without coupling.

2. Refraction law for acoustic energy streamlines

Consider linear acoustic waves in a three-dimensionally inhomogeneous, quiescent, inviscid fluid with piecewise continuous sound speed $c(\mathbf{r})$ and mass density $\rho(\mathbf{r})$. Here $\mathbf{r}$ is the position vector. Any surface Γ, at which at least one of the environmental parameters c and ρ is discontinuous, forms a fluid-fluid interface. To describe refraction of acoustic energy streamlines at an interface, it is sufficient to study transformation at the interface of the acoustic power-flux density $\mathbf{I}$, which determines the tangent to the energy streamlines. Acoustic power-flux density is given by the Umov vector[16,17]

$$\mathbf{I} = p\mathbf{v}, \tag{1}$$

where p and $\mathbf{v}$ are the acoustic pressure and fluid velocity in a sound wave.

Boundary conditions on a fluid-fluid interface[16,17] consist in continuity of the acoustic pressure and normal component of the oscillatory velocity:

$$p^{(+)} = p^{(-)}, \quad \mathbf{N}\cdot\mathbf{v}^{(+)} = \mathbf{N}\cdot\mathbf{v}^{(-)}, \quad \mathbf{r}\in\Gamma. \tag{2}$$

Here $\mathbf{N}$ is a unit normal to the interface, and superscripts "(+)" and "(−)" designate the limiting values a physical quantity takes when an observation point approaches a point on Γ from opposite sides of the interface. Since Eq. (2) holds at every point on Γ, derivatives of the acoustic pressure in the directions tangential to Γ are also continuous at the interface. Acoustic pressure gradient and fluid velocity in a linear sound wave are related by the linearized (with respect to the wave amplitude) Euler equation[17] $\partial\mathbf{v}/\partial t = -\nabla p/\rho$. From the Euler equation and continuity of the tangential to Γ components of ∇p, it follows that

$$\rho^{(+)}\,\partial\,\mathbf{v}_\tau^{(+)}/\partial t = \rho^{(-)}\,\partial\,\mathbf{v}_\tau^{(-)}/\partial t, \quad \mathbf{r}\in\Gamma, \tag{3}$$

where the vector $\mathbf{v}_\tau = \mathbf{v} - (\mathbf{v}\cdot\mathbf{N})\mathbf{N}$ has the meaning of the component of the fluid velocity in the directions tangential to the interface. For continuous waves and also for a rather broad class of transient waves, for which $\mathbf{v}\to 0$ at $t\to-\infty$ and/or $t\to+\infty$, from Eq. (3), we obtain

$$\rho^{(+)}\mathbf{v}_\tau^{(+)} = \rho^{(-)}\mathbf{v}_\tau^{(-)}, \quad \mathbf{r}\in\Gamma. \tag{4}$$

For the normal $I_n = \mathbf{N}\cdot\mathbf{I}$ and tangential $\mathbf{I}_\tau = \mathbf{I} - I_n\mathbf{N}$ components of the power-flux density at the interface, from the boundary conditions (2) and (4), we find that

$$I_n^{(+)} = I_n^{(-)}, \quad \rho^{(+)}\mathbf{I}_\tau^{(+)} = \rho^{(-)}\mathbf{I}_\tau^{(-)}, \quad \mathbf{r}\in\Gamma. \tag{5}$$

While it is well known that the normal component of acoustic power-flux density is continuous at interfaces and the tangential component is generally discontinuous,[16,17] the second equation in Eq. (5), which establishes the proportionality of tangential components of the power-flux density on opposite sides of the fluid-fluid interface, appears to be new.

For any point $\mathbf{r}_0$ on Γ, define the plane of incidence as the plane containing $\mathbf{r}_0$ and parallel to vectors $\mathbf{N}(\mathbf{r}_0)$ and $\mathbf{I}^{(+)}(\mathbf{r}_0)$. Equation (5) shows that the vector $\mathbf{I}^{(-)}(\mathbf{r}_0)$ is in the plane of incidence. Let ψ be the acute angle between vectors $\mathbf{I}$ and $\mathbf{N}$ at the interface, so that $\tan\psi = I_\tau/I_n$. From Eq. (5) we find the refraction law for acoustic energy streamlines:

$$\rho^{(+)}\tan\psi^{(+)} = \rho^{(-)}\tan\psi^{(-)}, \quad \mathbf{r}\in\Gamma. \tag{6}$$

This is Chapman's refraction law, which was originally derived in Ref. 10 from other considerations under restrictive assumptions of a plane interface of homogeneous fluids.

The above derivation of Eq. (6) applies to instantaneous values of power-flux density $\mathbf{I}$ and corresponding streamlines in acoustic fields with arbitrary time dependence. If complex notation is used for continuous waves, and P and $\mathbf{V}$ designate complex amplitudes of acoustic pressure $P(\mathbf{r})\exp(-i\omega t)$ and fluid velocity $\mathbf{V}(\mathbf{r})\exp(-i\omega t)$ in a wave of frequency ω, Eqs. (2) and (4) remain valid when P and $\mathbf{V}$ are substituted for p and $\mathbf{v}$. Hence, Eq. (5) holds also for the power-flux density[17] $\bar{\mathbf{I}}=0.5\,\mathrm{Re}(P*\mathbf{V})$ averaged over the wave period. Here "$*$" designates complex conjugation. Therefore, Chapman's refraction law (6) applies to both instantaneous and cycle-averaged power fluxes and corresponding energy streamlines.

At a perfectly reflecting boundary, the normal component I_n of vector $\mathbf{I}$ is zero due to energy conservation. From $\tan\psi=I_\tau/I_n$, one might surmise that energy streamlines are tangent to perfectly reflecting boundaries. However, this is not always the case. For a wave incident from the (+) side of a surface Γ, pressure-release and rigid boundaries Γ can be considered as limiting cases of a fluid-fluid interface of the same geometry, when $m\equiv\rho^{(-)}/\rho^{(+)}$ tends to zero and infinity, respectively. In the limit $m\to\infty$, it follows from Chapman's law (6) that $\psi^{(+)}=\pi/2$, and acoustic energy streamlines are indeed tangent to rigid boundaries, except in the case of normal incidence, where $\psi^{(+)}=\psi^{(-)}=0$. [Angle $\psi^{(+)}$ on a boundary is understood to be the limit of $\psi^{(+)}$ at a point in fluid away from the boundary when the distance from the point to the boundary tends to zero.] In the other case, when $m\to 0$, Chapman's law (6) gives $\psi^{(+)}=0$. Hence, energy streamlines are orthogonal to pressure-release boundaries. This behavior results from the tangential component I_τ of the power-flux density vanishing faster than its normal component I_n when the observation point approaches a pressure-release boundary. The above conclusions can be verified using explicit field expressions in the case of plane wave reflection from a plane boundary.

3. Refraction law for energy streamlines of electromagnetic waves

Consider electromagnetic waves in an inhomogeneous dielectric with piecewise continuous dielectric permittivity $\varepsilon(\mathbf{r})$ and magnetic permeability $\mu(\mathbf{r})$. We assume that the medium is invariant with respect to translations in a certain direction. Then interfaces, i.e., surfaces at which ε and/or μ are discontinuous, are cylindrical surfaces. Introduce orthogonal Cartesian coordinates x, y, and z with the y coordinate axis along the direction of translation invariance. Then ε and μ are independent of y, and any interface Γ can be specified by the equation $f(x,z)=0, -\infty<y<\infty$. The function f is assumed to be sufficiently smooth so that unit normal $\mathbf{N}$ to Γ exists at every point $\mathbf{r}\in\Gamma$. In particular, the medium can be a stratified medium, where $\varepsilon=\varepsilon(z)$ and $\mu=\mu(z)$. In this special case, the interfaces are planes $z=$const. It can be demonstrated that linearly polarized transverse electric (TE) and transverse magnetic (TM) waves propagate in such an environment without coupling.[16] In TE waves, also known as horizontally polarized waves, electric field vector $\mathbf{E}$ is orthogonal to the xz plane. In TM waves, also known as vertically polarized waves, magnetic field vector $\mathbf{H}$ is orthogonal to the xz plane. At an interface, electromagnetic fields satisfy the following boundary conditions:[18]

$$\mathbf{N}\times(\mathbf{E}^{(+)}-\mathbf{E}^{(-)})=0, \quad \mathbf{N}\times(\mathbf{H}^{(+)}-\mathbf{H}^{(-)})=0, \quad \mathbf{r}\in\Gamma, \tag{7}$$

$$\varepsilon^{(+)}\mathbf{N}\cdot\mathbf{E}^{(+)}=\varepsilon^{(-)}\mathbf{N}\cdot\mathbf{E}^{(-)}, \quad \mu^{(+)}\mathbf{N}\cdot\mathbf{H}^{(+)}=\mu^{(-)}\mathbf{N}\cdot\mathbf{H}^{(-)}, \quad \mathbf{r}\in\Gamma. \tag{8}$$

The power-flux density in the electromagnetic field is given by the Poynting vector[18]

$$\mathbf{S} = c(\mathbf{E} \times \mathbf{H})/4\pi, \tag{9}$$

where c is the speed of light in a vacuum. In terms of normal and tangential to the interface components of the Poynting and field vectors, Eq. (9) becomes

$$S_n = c\mathbf{N} \cdot (\mathbf{E}_\tau \times \mathbf{H}_\tau)/4\pi, \quad \mathbf{S}_\tau = (c/4\pi)[E_n(\mathbf{N} \times \mathbf{H}_\tau) - H_n(\mathbf{N} \times \mathbf{E}_\tau)], \quad \mathbf{r} \in \Gamma. \tag{10}$$

According to Eq. (7), vectors $\mathbf{E}_\tau$ and $\mathbf{H}_\tau$ are continuous at the interface. Hence, the normal component of the Poynting vector is also continuous at the interface. (Continuity of S_n is well known and can be viewed as a corollary of the energy conservation law.) The tangential component of the Poynting vector is generally discontinuous at the interface.

In the special case of TE waves in a translationally invariant medium, $E_n=0$. It then follows from Eq. (10) and the boundary condition (8) that

$$\mu^{(+)}\mathbf{S}_\tau^{(+)} = \mu^{(-)}\mathbf{S}_\tau^{(-)}, \quad \mathbf{r} \in \Gamma. \tag{11}$$

Similarly, taking into account that $H_n=0$, for TM waves one finds from Eq. (10) and the boundary condition (8) that

$$\varepsilon^{(+)}\mathbf{S}_\tau^{(+)} = \varepsilon^{(-)}\mathbf{S}_\tau^{(-)}, \quad \mathbf{r} \in \Gamma. \tag{12}$$

Thus, the products $\mu\mathbf{S}_\tau$ and $\varepsilon\mathbf{S}_\tau$ are continuous at the interface in TE and TM waves, respectively.

The Poynting vector has zero y component in TE and TM waves, and the xz plane serves as the plane of incidence. Let ψ be an acute angle between $\mathbf{S}$ and the normal $\mathbf{N}$ to the interface. Then $\tan\psi = S_\tau/S_n$. From Eqs. (11) and (12) and continuity of S_n, we find the refraction law for energy streamlines

$$\mu^{(+)} \tan\psi_{\mathrm{TE}}^{(+)} = \mu^{(-)} \tan\psi_{\mathrm{TE}}^{(-)}, \quad \varepsilon^{(+)} \tan\psi_{\mathrm{TM}}^{(+)} = \varepsilon^{(-)} \tan\psi_{\mathrm{TM}}^{(-)}, \quad \mathbf{r} \in \Gamma \tag{13}$$

of TE and TM waves, respectively. It should be emphasized that refraction laws are distinct for TE and TM waves (except at normal incidence, where TE and TM waves become indistinguishable). Therefore, no universal refraction law, which is independent of the wave polarization, can possibly exist for energy streamlines in a generic electromagnetic field.

The above considerations apply to fields with arbitrary time dependence in media without dispersion. For continuous waves, ε and μ can depend on wave frequency ω in addition to spatial coordinates but are still assumed to be real-valued. Refraction laws (13) remain valid for instantaneous power fluxes and for energy streamlines defined in terms of cycle-averaged power fluxes. Validity of the refraction laws in the latter case can be established by either noting that Eqs. (11) and (12) and $S_n^{(+)} = S_n^{(-)}$ hold at each moment of time or repeating the above analysis for the cycle-averaged power-flux density. For streamlines of the cycle-averaged energy, Eq. (13) was obtained earlier[13] in the special case of plane TM wave refraction in a system consisting of a homogeneous layer between homogeneous half-spaces. Refraction laws (13) for electromagnetic waves also could have been derived, at least in the case of a plane boundary of homogeneous media, from Chapman's refraction law (6) for sound by employing the analogy between electromagnetic and acoustic wave refraction. Brekhovskikh[16] demonstrated that results obtained for sound describe reflection and refraction of linearly polarized electromagnetic waves, if the mass density ratio $\rho^{(+)}/\rho^{(-)}$ is replaced by $\mu^{(+)}/\mu^{(-)}$ for TE waves and $\varepsilon^{(+)}/\varepsilon^{(-)}$ for TM waves.

4. Discussion

Chapman's refraction law (6) is apparently different from Snell's refraction law,[17,18] which describes refraction of plane waves at a plane interface of homogeneous media as well as refraction of rays at an interface in homogeneous or inhomogeneous media. Let $\mathbf{a}_i$, $\mathbf{a}_r$, and $\mathbf{a}_t$ designate vectors which are tangential, respectively, to incident, reflected, and refracted rays (or parallel to wave vectors of incident, reflected, and transmitted plane waves). Snell's law states that normal $\mathbf{N}$ to the interface and vectors $\mathbf{a}_i$, $\mathbf{a}_r$, and $\mathbf{a}_t$ at the point of reflection lie in the same

plane (plane of incidence) and that $\theta_i=\theta_r$, $n^{(+)}\sin\theta_i=n^{(-)}\sin\theta_t$, where θ_i, θ_r, and θ_t are acute angles between $\mathbf{N}$ and $\mathbf{a}_i$, $\mathbf{a}_r$, and $\mathbf{a}_t$. Here $n^{(+)}$ and $n^{(-)}$ are the values the refraction index $n(\mathbf{r})$ takes on the sides of the interface where incident and transmitted waves exist, respectively. Refraction index can be defined as $n(\mathbf{r})=\sqrt{\varepsilon(\mathbf{r})\mu(\mathbf{r})}$ for electromagnetic waves; for acoustic waves $n(\mathbf{r})=s_0/s(\mathbf{r})$, where s is sound speed and s_0 is the value of s at an arbitrary chosen fixed point. Snell's law expresses the equality of tangential to the interface components of wave vectors of incident, reflected, and transmitted waves or, equivalently, equality of phase speeds of traces of these three waves on the interface.[17,18]

It is easy to see that the incidence planes implied in Snell's and Chapman's laws coincide. However, the two laws describe wave refraction in terms of distinct material parameters and predict rather different dependences between incidence and refraction angles. In particular, for electromagnetic waves, Snell's law, unlike Chapman's law (13), is independent of wave polarization. In acoustics, the difference between Snell's and Chapman's laws becomes particularly stark when one considers interfaces with either $\rho^{(+)}=\rho^{(-)}$, $n^{(+)}\neq n^{(-)}$ or $\rho^{(+)}\neq\rho^{(-)}$, $n^{(+)}=n^{(-)}$. One may wonder how two refraction laws can possibly describe the same process.

To understand why Snell's and Chapman's laws are different, it is important to realize that incidence angles as defined from wave power-flux density are generally different from incidence angles defined in terms of wave vector or ray direction. Indeed, power-flux density is uniquely defined at any point in space and moment in time. On the other hand, several rays can pass through a given point simultaneously. The direction of the power flux and tangent to energy streamline are associated with the total wave field and do not necessarily coincide with directions of any of the rays. When a wave is reflected from an interface, wave power flux is a result of the interference of incident and reflected waves and, except for the cases of normal incidence or reflection coefficient being zero, is not parallel to the wave vector of the incident wave. While Snell's law describes the kinematics of wave reflection and refraction, direction of energy streamlines depends on amplitude of reflected wave. For electromagnetic waves, reflection coefficients are polarization-dependent.[16,18] That is why Chapman's law (13) takes different forms for TE and TM waves.

Another important distinction between Chapman's and Snell's laws lies in the difference between their applicability domains. The concept of energy streamlines applies to arbitrary wave fields, while Snell's law refers to wave fields of special, albeit extremely important, types, namely, to plane waves and to short-wave fields, which can be described in ray terms.

Refraction angles $\psi^{(-)}$ and θ_t coincide if the transmitted plane wave or refracted ray never returns to the interface as a result of reflection at other interfaces and boundaries or continuous refraction in the medium. Incidence angles $\psi^{(+)}$ and θ_i will also coincide when the reflection coefficient equals zero. Then angles θ_t and θ_i satisfy both Snell's and Chapman's laws. For the angle of intromission of a plane acoustic wave, i.e., the incidence angle at which there is no reflection from a fluid-fluid interface, from Eqs. (6) and Snell's law, we find that

$$\tan^2\theta_i=(n^2-m^2)/(1-n^2), \tag{14}$$

where $n\equiv n^{(-)}/n^{(+)}=s^{(+)}/s^{(-)}$ and $m=\rho^{(-)}/\rho^{(+)}$. The angle of intromission exists when either $1<n<m$ or $m<n<1$. Equation (14) agrees with results[17] obtained from an explicit expression for the plane wave reflection coefficient. From Snell's and Chapman's (13) laws for electromagnetic waves it follows that Eq. (14) gives the incidence angle at which the reflection coefficient is zero for TM or TE waves, when $n\equiv n^{(-)}/n^{(+)}=[\varepsilon^{(-)}\mu^{(-)}/\varepsilon^{(+)}\mu^{(+)}]^{1/2}$ and $m=\varepsilon^{(-)}/\varepsilon^{(+)}$ or $m=\mu^{(-)}/\mu^{(+)}$, respectively. In the case where $\mu^{(-)}=\mu^{(+)}$, Eq. (14) reduces to the known expression $\tan\theta_i=n$ for Brewster angle[16,18] at which there is no reflection of TM waves.

In conclusion we note that Chapman's law can be extended to other wave types. For instance, consider an isotropic elastic solid with piecewise continuous parameters, which are independent of Cartesian coordinate y. It can be demonstrated that in such a system horizontally polarized shear waves, in which particle displacement is parallel to the y coordinate axis and the field is independent of y, propagate without coupling to compressional waves and vertically polarized shear waves.[17] The refraction law for energy streamlines of horizontally polarized

shear waves is derived quite similarly to the acoustic refraction law (6) and differs from it only by replacing mass density ρ by the reciprocal value $1/\mu$ of the shear modulus $\mu(\mathbf{r})$ of the solid. It is not known whether simple refraction laws exist for energy streamlines of elastic wave fields of other polarizations in solids.

Acknowledgments

This work was supported in part by the Office of Naval Research through Grant No. N00014-08-1-0100. The author thanks D. M. F. Chapman for drawing attention to the problem of energy streamline refraction and for several stimulating discussions.

References and links

[1]L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, Course of Theoretical Physics Vol. **6** (Pergamon, New York, 1982), Chap. 1.
[2]N. A. Umov, *Selected Works* (GITTL, Moscow, Leningrad, 1950) (in Russian).
[3]R. V. Waterhouse, T. W. Yates, D. Feit, and Y. N. Liu, "Energy streamlines of a sound source," J. Acoust. Soc. Am. **78**, 758–762 (1985).
[4]C. J. Chapman, "Energy paths in edge waves," J. Fluid Mech. **426**, 135–154 (2001).
[5]R. V. Waterhouse and D. Feit, "Equal-energy streamlines," J. Acoust. Soc. Am. **80**, 681–684 (1986).
[6]R. V. Waterhouse, D. G. Crighton, and J. E. Ffowcs-Williams, "A criterion for an energy vortex in a sound field," J. Acoust. Soc. Am. **81**, 1323–1326 (1987).
[7]F. J. Fahy, *Sound Intensity*, 2nd ed. (E&FN Spon, London, 1995).
[8]V. A. Zhuravlev, I. K. Kobozev, and Yu. A. Kravtsov, "Detecting dislocations by measuring the energy flux of an acoustic field," Sov. Phys. JETP **77**, 808–814 (1993).
[9]C. F. Chien and R. V. Waterhouse, "Singular points of intensity streamlines in two-dimensional sound fields," J. Acoust. Soc. Am. **101**, 705–712 (1997).
[10]D. M. F. Chapman, "Using streamlines to visualize acoustic energy flow across boundaries," J. Acoust. Soc. Am. **124**, 48–56 (2008).
[11]V. A. Babeshko, E. V. Glushkov, and N. V. Glushkova, "Energy vortices and backward fluxes in elastic waveguides," Wave Motion **16**, 183–192 (1992).
[12]E. Glushkov and N. Glushkova, "Blocking property of energy vortices in elastic waveguides," J. Acoust. Soc. Am. **102**, 1356–1360 (1997).
[13]Z. M. Zhang and B. J. Lee, "Lateral shift in photon tunneling studied by the energy streamline method," Opt. Express **14**, 9963–9970 (2006).
[14]B. J. Lee, K. Park, and Z. M. Zhang, "Energy pathways in nanoscale thermal radiation," Appl. Phys. Lett. **91**, 153101 (2007).
[15]T. L. Foreman, "An exact ray theoretical formulation of the Helmholtz equation," J. Acoust. Soc. Am. **86**, 234–246 (1989).
[16]L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960), pp. 6–28, 114, and 170.
[17]L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media I: Plane and Quasi-Plane Waves* (Springer-Verlag, Berlin, 1990), pp. 2–22 and 117.
[18]M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed. (Cambridge University Press, Cambridge, 2001), pp. 5–9 and 40–45.
[19]E. A. Skelton and R. V. Waterhouse, "Energy streamlines for a spherical shell scattering plane waves," J. Acoust. Soc. Am. **80**, 1473–1478 (1986).
[20]J. Zhang and G. Zhang, "Analysis of acoustic radiation and scattering from a submerged spherical shell by energy streamlines," J. Acoust. Soc. Am. **88**, 1981–1985 (1990).
[21]X. D. Xu, H. P. Lee, C. Lu, and J. Y. Guo, "Streamline representation for structural intensity fields," J. Sound Vib. **280**, 449–454 (2005).
[22]E. V. Glushkov, N. V. Glushkova, and M. V. Golub, "Blocking of traveling waves and energy localization due to the elastodynamic diffraction by a crack," Acoust. Phys. **52**, 259–269 (2006).

In situ measurement of permeability

Lin Lin and M. L. Peterson
Department of Mechanical Engineering, University of Maine, Orono, Maine 04469
lin.lin@maine.edu, michael.peterson@maine.edu

Alan R. Greenberg
Membrane Applied Science and Technology Center, University of Colorado, Boulder, Colorado 80309
alan.greenberg@colorado.edu

Benjamin A. McCool
ExxonMobil Research and Engineering Company, Annandale, New Jersey 08801
benjamin.a.mccool@exxonmobil.com

Abstract: Permeability of a porous material with a fluid interface is shown to be related to the propagation of the slow longitudinal wave. The propagation threshold of the slow longitudinal wave occurs when the wave number is higher than the critical wave number, k_{cr}. Measuring k_{cr} can provide the intrinsic permeability of a porous interface. The ability to detect a change in k_{cr} due to differences in pore size and fluid viscosity is demonstrated. This measurement is demonstrated in a model material with two pore sizes and two fluid viscosities. The results suggest opportunities to extend the method to a range of materials of scientific and technical importance.

PACS numbers: 43.20.Gp, 43.20.Ye, 43.35.Yb [AN]
Date Received: October 16, 2008 **Date Accepted:** January 8, 2009

1. Introduction

In 1956 Biot[1] predicted that fluid-saturated porous materials are able to support the propagation of two longitudinal waves, the fast and slow waves. The slow wave's velocity is always lower than both the compressional wave velocity in the fluid and the longitudinal velocity in the solid frame. The slow longitudinal wave, which does not exist in dense solids, represents an out-of-phase movement of the pore fluid relative to the matrix. This motion is very sensitive to the viscosity of the fluid and the permeability of the porous formation.[2] This slow wave is important since porous materials are ubiquitous in nature and technology in the form of fibrous and granular materials, sediments, bones, and oil-bearing rocks. The slow wave can thus serve an important function as a tool for investigating the ability of fluids to penetrate into these materials.

Experimental observation of Biot's slow wave did not occur for nearly 30 years after Biot's work until Plona[3] observed the slow wave in a specimen made from sintered glass beads. Following Plona's work, extensive consideration was given to the slow wave in different porous media, especially geophysical materials[4,5] and human bones.[6–8] However, most studies of the slow wave were limited to considering the wave as a complication of elastic wave theory when investigating porous materials. Only recently has the possibility been raised that a useful relationship exists between the slow wave and material permeability.[9]

A significant difficulty in studying or utilizing slow waves is that the slow wave is not readily detectable. A particularly important description of the interaction of the slow wave with a porous interface is contained in the critical wave number, k_{cr}. Below the critical wave number the slow wave is highly attenuated since the skin-depth of the fluid in the pores is large relative to the pore size. Ideally, k_{cr} represents a bifurcation below which the slow wave does not propagate.[10] Because of the dependence of the slow wave on the motion of the fluid, the detection of the critical wave number provides a basis for measuring the dynamic interaction between the matrix and the pore fluid.

In scientific and technical applications of porous materials, permeability often represents the reason the porous structure is of interest. For example, when a filtration membrane is used for water treatment or chemical separations, the properties of the porous structure such as pore size, porosity, and tortuosity are constant. However, the permeability typically will change during membrane operation.[11] Permeability is normally determined by measuring the fluid pressure difference across a porous specimen, and the resulting flow rate through the material. There are various existing techniques that utilized this basic scheme.[12] However, many of these methods are invasive and some use potentially dangerous materials. In general, permeability cannot be directly measured.[13] The measurement reported in this paper is the first example of using the slow wave transition to quantify permeability at an interface in real time.

2. Exact solution of critical wave number

The problem of waves in porous materials has been considered by a number of authors.[14,15] Following the notation of Edelman,[10] the bulk wave propagated in an infinite fluid-saturated porous medium with a porosity of n_0 has an exact solution for the critical wave number, k_{cr}, of

$$k_{\mathrm{cr}} = y\frac{\pi}{{\rho_0}^S {U_{II}}^S}, \tag{1}$$

where

$$y = y({\rho_0}^S, {\rho_0}^F, {U_{II}}^S, U^F). \tag{2}$$

${\rho_0}^S$ and ${\rho_0}^F$ represent the partial mass densities of the solid and fluid, respectively. ${U_{II}}^S$ is the velocity of the longitudinal wave in the skeleton, and U^F is the velocity of the fluid. In addition, π is related to the porosity n_0, viscosity of the fluid μ^F, and permeability of the porous medium κ by $\pi = n_0\mu^F/\kappa$. If $r = {\rho_0}^F/{\rho_0}^S$ and $c_f = U^F/{U_{II}}^S$, then y can be described as

$$y^2 = \sqrt[3]{-\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}} + \sqrt[3]{-\frac{b}{2} - \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}} - \frac{B}{3A}, \tag{3}$$

where

$$a = \frac{1}{3}\left(3\left(\frac{C}{A}\right) - \left(\frac{B}{A}\right)^2\right), \tag{4}$$

$$b = \frac{1}{27}\left(2\left(\frac{B}{A}\right)^3 - 9\left(\frac{B}{A}\right)\left(\frac{C}{A}\right) + 27\left(\frac{D}{A}\right)\right), \tag{5}$$

$$A = 4c_f^2(c_f - 1)^4(c_f + 1)^4, \tag{6}$$

$$B = -(c_f - 1)^2(c_f + 1)^2\left[\left(c_f^2 + \frac{1}{r^2}\right)(c_f^4 - 10c_f^2 + 1) + 24c_f^2\frac{c_f^2 + 1}{r}\right], \tag{7}$$

$$C = -2\left(c_f^4 + \frac{1}{r^4}\right)(c_f^4 - 4c_f^2 + 1) + \left(\frac{c_f^2}{r} + \frac{1}{r^3}\right)(5c_f^6 + 3c_f^4 + 3c_f^2 + 5) + \frac{1}{4}\left(\frac{c_f^8 + 108c_f^6 - 122c_f^4 + 108c_f^2 + 1}{r^2}\right), \tag{8}$$

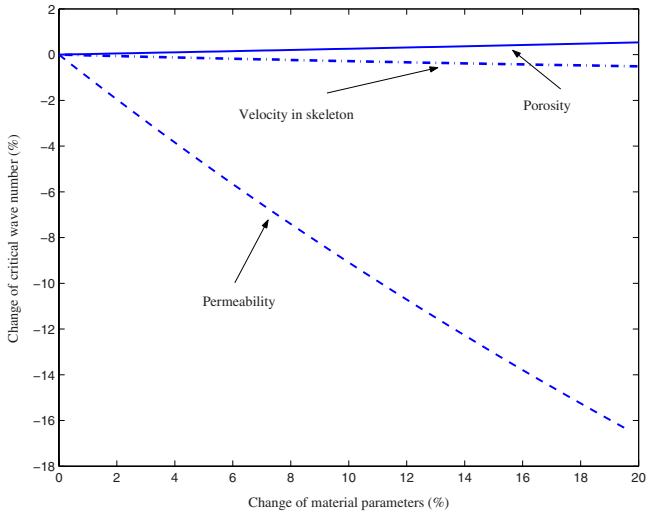


Fig. 1. (Color online) The critical wave number is sensitive to changes in material properties. The solid line shows the increase in the critical wave number when the total porosity is increased. The dash-dotted line shows the effect of changes in the longitudinal velocity of the skeleton. These changes are modest in comparison to those due to a change in the permeability (dashed line).

$$D = -(r+1)^3 \frac{(c_f^2 r + 1)^3}{r^6}. \tag{9}$$

This exact solution of the critical wave number provides a convenient form to directly relate the critical wave number measurement to permeability. If the goal is to configure experiments, this simple solution can be used to identify the range of frequencies for detecting the critical wave number in porous materials.

The critical wave number has been determined for a wide range of porous materials and is, for example, on the order of 70 m^{-1} for sandstone and 0.02 m^{-1} for human cancellous bone. The sensitivity of k_{cr} to material parameters can be determined by varying a single parameter with other properties held constant, as shown in Fig. 1. In many porous materials such as bone, the total porosity will have a large spatial variation. However, when the porosity increases by 20%, the value of k_{cr} increases by only 0.53%. The velocity of the longitudinal wave in the skeleton may also be uncertain for some materials such as partially cross linked polymers used in membranes for separations. Similarly, a 20% increase in the longitudinal wave velocity in the skeleton results in a 0.51% decrease in k_{cr}. In contrast, k_{cr} decreases by approximately 16.7% when the permeability increases by 20%. Thus, while the critical wave number is relatively insensitive to the parameters for which significant uncertainty exists, k_{cr} is highly sensitive to permeability in a porous material. Consequently, a technique that enables k_{cr} to be accurately measured makes possible the determination of the interface permeability, even when many other characteristics of the permeable interface may be uncertain or variable.

3. Method

The slow wave is not readily detectable since the wave is evanescent when a viscous fluid fills the pores.[2] In a through transmission configuration, either the sample must be thick enough for the waves to be separated in time or the existence of the slow wave must be determined from an interference pattern. The ability to access both sides of a sample is also not possible in many applications. However, in a reflection from an interface, energy is transferred into the slow wave that should be detectable as a decrease in the reflection coefficient. Wu *et al.*[16] developed a theoretical analysis to study the reflection and transmission of elastic waves at the interface

Table 1. Material properties of experimental samples.

	Alumina sample 1	Alumina sample 2
Porosity	66%	57%
Particle size (μm)	0.3	1
Density of alumina powder (kg/m^3)	3960	3960
Bulk modulus of solid (N/m^2)	228×10^9	228×10^9
Bulk modulus of frame (N/m^2)	26×10^9	42×10^9
Shear modulus of the frame (N/m^2)	17×10^9	28×10^9
Tortuosity	1.4	1.8
Permeability (m^2)	1.6×10^{15}	7.3×10^{15}

between a fluid and a porous solid. At wave numbers above the critical wave number the reflection coefficient should decrease significantly. This model system builds a bridge between the exact solution of the critical wave number and the experimental permeability measurement. Previous experimental results have not considered the change in reflection in spite of its potential importance as a measurement method.

To provide an appropriate permeability as an input for the modeling, a falling head permeability measurement was used on the samples.[17] The permeability, κ, can be determined based on Darcy's law of fluid flow through porous media according to the following formula:

$$\kappa = \frac{aL}{At}\ln\left(\frac{H_0}{H_1}\right)\frac{\mu}{\rho g}, \tag{10}$$

where a and A are cross-sections of the measuring tube and sample, L denotes the sample thickness, t is the time for the water head to fall from H_0 to H_1, μ is the viscosity of the fluid, ρ is the density of the fluid, and g is the acceleration due to gravity.

4. Acoustic experiment

To experimentally test this hypothesis using two different specimens, alumina porous samples were prepared using 1 and 0.3 μm alumina powders. The powder was placed in a graphite die and hot-pressing at 1375 °C for 48 min in an inert argon atmosphere. The material properties of the two samples are shown in Table 1. Using Eq. (1), the cut-off frequencies corresponding to the calculated critical wave number k_{cr} for the two samples are 10 MHz (0.3 μm; 66% total porosity) and 1 MHz (1 μm; 57% total porosity), respectively.

The test procedure was based on a narrow-band ultrasonic methodology. The equipment includes a function/arbitrary waveform generator (Agilent 33250A, Santa Clara, CA), a digital oscilloscope (Tektronix TDS 520A, Wilsonville, OR), and a gated rf amplifier (Ritec model GA-2500A, Warwick, RI). The samples were vacuum impregnated with fluid. Six broadband ultrasonic transducers (0.5, 1, 2.25, 5, 7.5, and 10 MHz) were excited with a narrow-band gated sine wave to produce measurements at frequencies from 400 kHz to 10 MHz. The angle of the sample was adjusted to provide maximum reflection in order to ensure a normally incident ultrasonic wave insonified the sample. The reflected signal was acquired using the same ultrasonic transducer used to insonify the sample. Three data points were taken at different locations for each sample at each frequency. Because of the homogeneous nature of the sample, no significant difference in the values was measured. For calibration of the apparatus a polished stainless steel reference was used. Figure 2 shows experimental data with the exact solution plotted against a calculated reflection curve for two different samples using water as the pore fluid. The two reflection coefficient curves are clearly separated in frequency domain. The critical wave number is also sensitive to fluid density, wave speed, and viscosity. In some practical applications the viscosity will change significantly during operation and thus impact permeability. Figure 3 compares the effect of using a lower viscosity fluid, acetone (viscosity 0.3 cP) with the

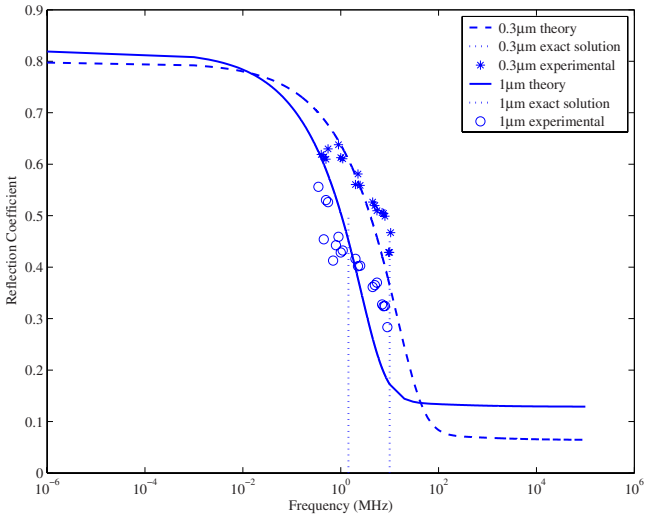


Fig. 2. (Color online) Reflection coefficient vs frequency in water for samples fabricated from two different alumina particle sizes. The dashed line and data point (∗) indicate the calculated and experimental reflection coefficients, respectively, for the 0.3 μm particle-size sample. The solid line and open data points (○) indicate the calculated and experimental reflection coefficients, respectively, for the 1 μm particle-size sample. The vertical dashed lines represent an exact solution for the critical frequencies for each sample.

same test using water (viscosity 1 cP). These data suggest that lower viscosity causes the reflection coefficient transition curve to shift to a lower frequency, which is consistent with the theory.

5. Conclusion

It is evident that this technique is able to sense a permeability change caused by either the difference in the pore structure or a change in the viscosity of the pore fluid. The trends in the

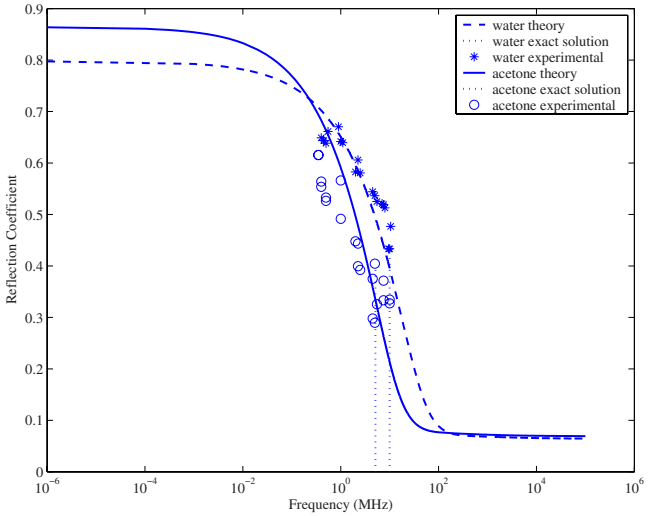


Fig. 3. (Color online) Reflection coefficients for the 0.3 μm sample in water and acetone. The dashed line and data points (*) indicate the calculated and experimental reflection coefficients, respectively, in water. The solid line and open data points (○) indicate the calculated and experimental reflection coefficients, respectively, in acetone. The vertical dashed lines represent an exact solution for the critical frequencies and indicate that a change in fluid viscosity of approximately three times produces a corresponding change in the critical frequency of approximately two times.

experiment are consistent with theory. Therefore, a technique based on the critical wave number measurement may be appropriate for estimating the permeability of a porous material. An appropriate inversion technique for measuring *in situ* permeability will open up a wide range of applications. One application example is the relationship between bone remodeling and interstitial fluid flow that can be probed using slow waves.[6] Many other examples exist for which the permeability controls dynamic processes in technological and natural systems.

Acknowledgments

The authors acknowledge support from the NSF Industry/University Cooperative Research Center for Membrane Applied Science and Technology at the University of Colorado at Boulder. They are also grateful for the valuable discussions and technical support from Dr. Thomas J. Plona (Schlumberger-Doll Research).

References and links

[1]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency ranges," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[2]P. B. Nagy, "Slow wave propagation in air-filled permeable solids," J. Acoust. Soc. Am. **93**, 3224–3234 (1993).
[3]T. J. Plona, "Observation of a second bulk compressional wave in a porous medium at ultrasonic frequencies," Appl. Phys. Lett. **36**, 259–261 (1980).
[4]F. A. Boyle and N. P. Chotiros, "Experimental detection of a slow acoustic wave in sediment at shallow grazing angles," J. Acoust. Soc. Am. **91**, 2615–2619 (1992).
[5]O. Kelder and D. M. Smeulders, "Observation of the Biot slow wave in water-saturated Nivelsteiner sandstone," Geophysics **62**, 1794–1796 (1997).
[6]R. Lakes, H. S. Yoon, and J. L. Katz, "Slow compressional wave propagation in wet human and bovine cortical bone," Science **220**, 513–515 (1983).
[7]A. Hosokawa and T. Otani, "Acoustic anisotropy in bovine cancellous bone," J. Acoust. Soc. Am. **103**, 2718–2722 (1998).
[8]Z. E. A. Fellah, J. Y. Chapelon, S. Berger, W. Lauriks, and C. Depollier, "Ultrasonic wave propagation in human cancellous bone: Application of Biot theory," J. Acoust. Soc. Am. **116**, 61–73 (2004).
[9]Z. E. A. Fellah, M. Fellah, F. G. Mitri, N. Sebaa, C. Depollier, and W. Lauriks, "Measuring permeability of porous materials at low frequency range via acoustic transmitted waves," Rev. Sci. Instrum. **78**, 114902 (2007).
[10]I. Edelman, "Bifurcation of the Biot slow wave in a porous medium," J. Acoust. Soc. Am. **114**, 90–97 (2003).
[11]M. Marcel, *Basic Principles of Membrane Technology* (Kluwer Academic, Dordrecht, 1991).
[12]R. Meyer and F. F. Krause, "Experimental evidence for permeability minima at low-velocity gas flow through naturally formed porous media," J. Porous Media **1**, 93–106 (1998).
[13]M. Sahimi, "Flow phenomena in rocks: From continuum models to fractals, percolation, cellular automata, and simulated annealing," Rev. Mod. Phys. **65**, 1393–1534 (1993).
[14]G. Mavko, T. Mukerji, and J. Dvorkin, *The Rock Physics Handbook* (Cambridge University Press, Cambridge, 1998).
[15]R. D. Stoll, *Sediment Acoustics*, Lecture Notes in Earth Sciences (Springer-Verlag, Berlin, 1989).
[16]K. Wu, Q. Xue, and L. Adler, "Reflection and transmission of elastic waves from a fluid-saturated porous solid boundary," J. Acoust. Soc. Am. **87**, 2349–2458 (1990).
[17]R. A. Freeze and J. A. Cherry, *Groundwater* (Prentice-Hall, Englewood Cliffs, NJ, 1979).

Plate wave stop-bands in periodically poled lithium niobate

I. V. Ostrovskii, V. A. Klymko, and A. B. Nadtochiy
Department of Physics and Astronomy, University of Mississippi, University, Mississippi 38677
iostrov@phy.olemiss.edu, vick@olemiss.edu, nadt@phy.olemiss.edu

Abstract: The dispersion curves of four lowest plate acoustic waves (PAWs) in the *ZX* cut of a periodically poled lithium niobate (PPLN) wafer are computed numerically and investigated experimentally. Experiment is in agreement with simulated dispersion curves. Calculations and measurements reveal the stop-bands in the dispersion curves of the PAW modes in PPLN despite a wafer that has uniform mechanical properties. At a specific wave number, within the frequencies of the stop-bands, the acoustic modes do not propagate.

PACS numbers: 43.20.Mv, 43.35.Cg [JM]
Date Received: December 24, 2008 **Date Accepted:** January 12, 2009

1. Introduction

Interest in the propagation of acoustic waves in two-dimensional and periodic structures has grown in recent years because of the intriguing properties of acoustic superlattices.[1,2] Existence of acoustic stop-bands or forbidden bands has been shown for a number of periodic media in different solid-based composite structures including one-dimensional ferroelectrics and periodically microtapered optical fibers.[1–6] At a certain wave number, within the frequencies of an acoustic stop-band, a specific acoustic mode cannot propagate. In the case of band gap that is sometimes refered to as a full or complete band gap, no acoustic phonon can be transmitted over a range of wave numbers.[6] The acoustic stop-bands for bulk and surface waves were discovered in two-dimensional composites based on $LiNbO_3$ with air holes,[6] piezoelectric $Bi_{12}GeO_{20}$ filler in a silica substrate,[7] and piezocomposites consisting of the layers of ZnO, PZT4, PZT5, CdS, and ZnO.[8] The main property of the superlattices is a periodicity in material physical of the composite structures including elastic, dielectric and piezoelectric constants and density. As a rule, the speed of an acoustic wave is different in the individual structure components. Despite extensive efforts in this field, the dispersion of plate acoustic waves (PAWs) in periodically poled lithium niobate (PPLN) wafers has not yet been investigated, in spite of contemporary interest in the phenomenology and applications of PPLN crystals.

In this work, we investigate the spectrum of PAWs propagating along *X* axis in the *Z*-cut of a PPLN wafer. PPLN is known to be photonic crystal, and our aim in this study is to search for acoustic stop-bands, thus establishing it as both a photonic and a phononic solid. The hypothesis that PPLN may have an acoustic stop-band is based on the effect of domain resonance observed in periodically polled lithium tantalite.[9]

2. Methods

A model of a PPLN wafer is shown in the inset of Fig. 1(a). The inversely directed arrows represent a periodic polarization in a multidomain array. For theoretical calculations, the inversely poled domains are introduced by periodically changing the sign of a piezoelectric constant in the equations of motion (1) and electrodynamics (2), $(-1)^{n-1}\cdot e_{m,ij}$,

$$\rho\frac{\partial^2 u_i}{\partial t^2}=\frac{\partial T_{ij}}{\partial x_j}=\frac{\partial}{\partial x_j}\left[c^E_{ijkl}\frac{\partial u_k}{\partial x_l}+(-1)^{n-1}\cdot e_{m,ij}\frac{\partial \varphi}{\partial x_m}\right], \quad (1)$$

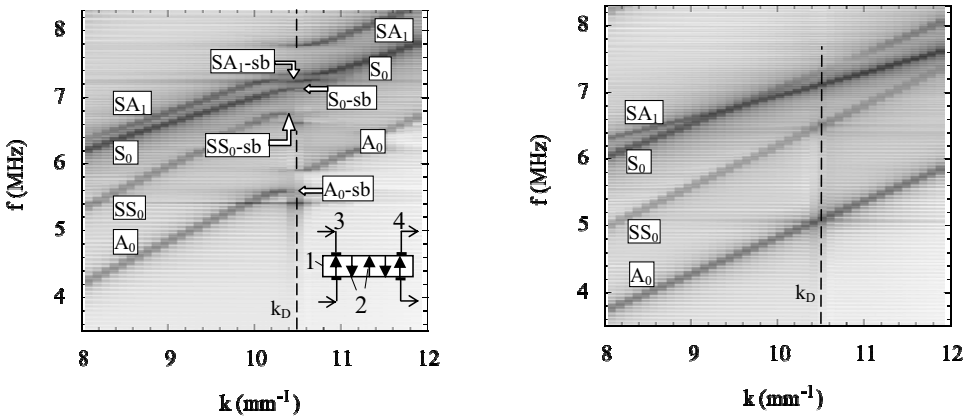


Fig. 1. Numerical dispersion curves for X-propagating PAW in the 0.5-mm-thick Z-cut PPLN with $N=216$ domains of 300-μm-width each: (a) FEM computation with regular piezoelectric constants of $LiNbO_3$. inset in panel (a): 1, PPLN wafer; 2, inversely poled domains; 3, input transducer; 4, output pickup. (b) FEM computation with ten-time-reduced piezoelectric constants. Vertical dashed lines show $k=k_D=(\pi/d)=10.47$ mm^{-1}. The same plate modes are observed in both panels (a) and (b), but the stop-bands appear only in panel (a).

$$\frac{\partial D_m}{\partial x_m}=\frac{\partial}{\partial x_m}\left[(-1)^{n-1}\cdot e_{m,ij}\frac{\partial u_i}{\partial x_j}+\varepsilon^S_{mj}\frac{\partial \varphi}{\partial x_j}\right]=0, \tag{2}$$

where $n=1,2,\ldots,N$ is the domain's number and N is the total number of the domains in the array; indices i, j, k, l, and m change from 1 to 3 representing the X, Y, and Z crystallographic directions; ρ is crystal density; u_i is the acoustic displacement in PAW; T_{ij} is the stress tensor; c^E_{ijkl} is the tensor of elastic moduli at constant electric field; ε^S_{ij} is the dielectric permittivity at constant strain; D_i is the electric displacement in PPLN; and φ is the electric potential. A superscript n standing for the domain number can change from $n=1$ for the first domain at $X=0$ to $n=N$ for the last domain at right end of the sample [see Fig. 1(a)]. In addition to Eqs. (1) and (2), the boundary conditions of zero acoustic stress and electric charge at the plate surfaces are applied. To solve Eqs. (1) and (2) along with the appropriate boundary conditions, one can use Hamilton's variation principle extended to a piezoelectric crystal, where the functional of the total work-energy is minimized to zero,[10]

$$\delta\int (E_{\rm kin}-E_{st}+E_d+W)dt=0. \tag{3}$$

For a wafer of unit width along the Y- axis, the term $E_{\rm kin}=\frac{1}{2}\int \dot{u}^2 dx_1 dx_3$ is the PAW's kinetic energy, the term $E_{st}=\frac{1}{2}\int \mathbf{S}^t\cdot\mathbf{T}dx_1dx_3$ is the PAW's elastic energy due to deformation $\mathbf{S}^t$, and the term $E_d=-\frac{1}{2}\int D_m\cdot(\partial\varphi/\partial x_m)dx_1dx_3$ is the contribution to the total energy due to dielectric displacement D_m that is defined in the parentheses in Eq. (2). The term W is the work done by an excitation force to generate the PAW. Furthermore, we have used the well-known finite element method (FEM) technique to compute the spectrum of PAWs in PPLN. In Fig. 1(a), a fragment of this spectrum near the stop-bands is shown for the four lowest modes. For lower frequencies ($f<3.5$ MHz) and wave vectors ($k<8$ mm^{-1}), the PAW dispersion curves are smooth lines like for regular modes in an acoustic waveguide. In Fig. 1(b), the same dispersion curves are calculated with ten times smaller piezoelectric coefficients, i.e., with $0.1\cdot(-1)^{n-1}\cdot e_{m,ij}$ instead of $(-1)^{n-1}\cdot e_{m,ij}$ in Eqs. (1) and (2).

3. Results and discussion

The dispersion curves of Fig. 1(a) are calculated for a PPLN sample with $N=216$ domains of width $d=300$ μm each. Four computed plate waves are the zero asymmetrical quasi-Lamb mode A_0, the zero shear symmetrical mode SS_0, the zero symmetrical quasi-Lamb wave S_0, and the first shear asymmetrical mode SA_1. The dispersion curves of all four modes (A_0, SS_0, S_0, and SA_1) are discontinued at wave number $k=10.47$ mm^{-1}, which is denoted in the figure by a vertical dashed line. This magnitude of k exactly corresponds to the domain resonance,[9] which takes place at $k_D=(\pi/d)$, where d is the domain width. Four stop-bands for the modes A_0, SS_0, S_0, and SA_1 are shown in Fig. 1(a) by arrows with the corresponding mode indicated. The stop-bands occur when the PAW wavelength $\lambda=2d$ and, consequently, $k=k_D=(\pi/d)$. The PAW modes cannot propagate within their stop-bands because there are no solutions for the acoustic displacements. To find out the physical origin of the stop-bands in Fig. 1(a), we performed a FEM-simulation with ten-times-smaller piezoelectric coefficients, the results of which are shown in Fig. 1(b). A comparison of Figs. 1(a) and 1(b) leads one to conclude that the existence of the stop-bands is a result of the piezoelectric interaction between PAW modes and periodically poled domains. To check this conclusion, we made the additional simulations with different domain widths. Calculations yielded a low frequency shift in stop-band position if the domain width d increased and a high frequency shift when the width decreased. This conclusion is also consistent with the piezoelectric coupling of the waves. Our estimates yield different magnitudes for the effective piezoelectric coupling coefficients K^2 characterizing PAW modes. For instance, at $k=k_D$ for observed plate waves, we have $K^2(SA_1)>K^2(A_0)>K^2(S_0)$. In Fig. 1(a), the widths of the stop-band, Δf, have the same inequality, that is, $\Delta f(SA_1)>\Delta f(A_0)>\Delta f(S_0)$, all at $k=10.47$ mm^{-1}. The mode SS_0 is not reliably computed nor measured at $k>10.47$ mm^{-1}, which means it is not transmitted through the multi-domain array. These acoustic stop-bands are quite unusual because they appear in a mechanically uniform medium having the same density and elastic moduli in all the domains. The only difference is the inverse polarization of neighboring domains. The physics behind these uncommon stop-bands may be attributed to a strong altering of the energies $\boldsymbol{E}_{st}$ and $\boldsymbol{E}_d$ at the acousto-domain resonance condition of $\lambda=2d$, which makes it impossible to satisfy Eq. (3).

The experimental setup for the excitation and detection of PAWs in PPLN is a standard computer controlled system with a rf generator to excite PAWs, small ultrasonic receiver that is connected to a digital oscilloscope, and a computer controlled stage that moves the receiver over PPLN sample surface. A sample [see Fig. 1(a) inset] is fabricated by applying a 22 kV/mm electric field along the Z axis to a 0.5-mm-thick optical grade LN wafer. The inverted ferroelectric domains are 300 μm in width, and thus their spatial period is twice of the domain width. The PAW is excited by applying rf voltage to the 0.9-mm-wide input electrodes that are deposited on the top and bottom surfaces of the PPLN wafer. The sample is placed on a support, which is attached to a computer controlled motorized stage. In a typical run, the stage makes 800 steps to cover the distance of 40.64 mm. Detector 4 is a small piezoelectric transducer brought into contact with the sample in a narrow spot of $\sim 0.15\times 0.2$ mm^2, which is smaller than domain width. Frequency performance of the receiver in the frequency range of the stop-bands, 5.4–7.7 MHz, is a monotonically increasing line without any local extrema. A signal from detector 4 [item 4 in Fig. 1(a)] goes to a digital oscilloscope that records a waveform $u_z(t_j)$ taken at the positions x_i after each single step. When the experimental run is completed, the data are stored in a two-dimensional array, $u_z(t_j,x_i)$. To obtain experimental dispersion curves, the double discrete Fourier transform converts the $u_z(t_j,x_i)$ matrix into its Fourier image $u_z(f_j,k_i)$. A gray-scale plot of the $u_z(f_j,k_i)$ matrix is the experimental spectrum presented in Fig. 2(a). The dispersion curves of Fig. 2(a) are consistent with the FEM computation of Fig. 1(a) for all modes. However, the amplitudes "indicated by the blacker colors in the figures" are different for the various modes. In particular, the modes A_0, S_0, and SA_1 are reliably detected. In contrary, the

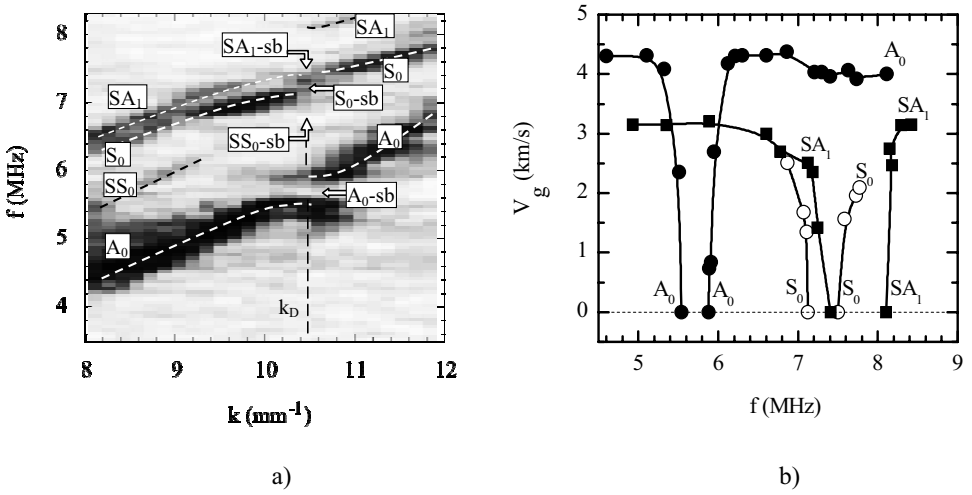


Fig. 2. (a) Experimental dispersion curves and (b) dispersion of the group velocities for *X*-propagating PAW in the 0.5-mm-thick *Z*-cut PPLN with $N=216$ domains of 300-μm-width each. The stop-bands in (a) panel appear at $k=k_D=(\pi/d)=10.47$ mm^{-1}. Arrows with sb-mark show the low frequency edges of the stop-bands. The dashed lines above the original computer generated graphs are plotted over the peak values of the experimental Fourier image $u_z(f_j,k_i)$. In panel (b), the zero group velocity points $V_g=0$ are consistent with the stop-band limits in corresponding modes.

mode SS_0, having mainly a shear displacement u_y, is weakly detected by the receiving transducer, probably because it mainly reads the vertical component of the total acoustical displacement.

As a final test, we analyze the group speeds V_g of three reliably detected modes, A_0, S_0, and SA_1. The magnitudes of V_g are calculated as $2\pi(df/dk)$ taken from the set of experimental data, which are graphically presented in Fig. 2(a). The results are given in Fig. 2(b). A horizontal dashed line represents the zero-speed limit $V_g=0$, where the mode is stopped. In Figs. 2(a) and 2(b), the stop-bands are detected in the frequency ranges of 5.60–5.90 MHz for A_0 mode, 7.10–7.45 MHz for S_0 mode, and 7.35–8.05 MHz for SA_1 mode. The low limits of the stop-bands are shown in Fig. 2(a) by the arrows corresponding to each mode, where the letters "sb" stand for "experimental stop-band." The SS_0 mode has mainly shear displacements along the *y* axis, and so it is not reliably detected. Nevertheless, a low limit of its stop-band may be read on a sufficiently resolved graph as 6.7 MHz, which is shown by dashed arrow in Fig. 2. The experimental stop-bands, positions, and widths are in agreement with the theoretical predictions of Fig. 1(a). The calculated stop-bands run from 5.60 to 5.90 MHz for A_0, 7.15 to 7.35 MHz for S_0, and 7.25 to 7.80 MHz for the SA_1 mode. The calculated stop-band of SS_0 mode starts at 6.75 MHz, as shown in Fig. 1(a).

4. Conclusions

In both the experiment and theory, all stop-bands appear at $k=10.47$ mm^{-1}, which unequivocally supports the hypothesis that the acousto-domain interaction is the origin of the stop-band effect. This conclusion is strongly supported by the additional simulations with different domain widths. In all cases, the stop-band effect takes place under the acousto-domain resonance condition of $\lambda=2d$, when the wavelength of any PAW is equal to two domain widths. The frequency width Δf of the stop-band depends on the piezoelectric coupling of an acoustic mode; for higher coupling coefficients, the stop-band width is larger. A piezoelectric field due to acoustic deformation is mainly responsible for the acousto-domain coupling, as it is clear from the comparison of Figs. 1(a) and 1(b). The results presented in this paper may stimulate research

toward new acousto-optic phenomena in phononic-photonic materials and lead to new applications in acousto-optics, acousto-electronics, ferroelectrics, and ultrasonics such as the construction of efficient microtransducers.

Acknowledgments

This research in part was supported by the research Grant No. FRP-2008, UM, and Graduate School Summer Research grant, UM. We are thankful to Dr. Mack Breazeale, Dr. Nico Declercq, and Dr. Michael McPherson for useful discussions.

References and links

[1]J. O. Vasseur, P. A. Deymier, B. Djafari-Rouhani, Y. Pennec, and A.-C. Hladky-Hennion, "Absolute forbidden bands and waveguiding in two-dimensional phononic crystal plates," Phys. Rev. B **77**, 085415 (2008).
[2]J. M. Vasseur, A.-C. Hladky-Hennion, B. Dubus, B. Djafari-Rouhani, and B. Morvan, "Design and characterization of stop-band filters using PZT layer on silicon substrate phononic crystals," J. Acoust. Soc. Am. **123**, 3039 (2008).
[3]M. A. Hawwa, "Acoustic/elastic stop-band interaction in waveguides involving two periodicities," J. Acoust. Soc. Am. **102**, 137–142 (1997).
[4]A. L. Shuvalov and A. S. Gorkunova, "Transverse acoustic waves in piezoelectric and ferroelectric antiphase superlattices," Phys. Rev. B **59**, 9070–9077 (1999).
[5]A. Diez, G. Kakarantzas, T. A. Birks, and P. St. J. Russell, "Acoustic stop-bands in periodically microtapered optical fibers," Appl. Phys. Lett. **76**, 3481–3483 (2000).
[6]S. Benchabane, A. Khelif, J.-Y. Rauch, L. Robert, and V. Laude, "Evidence for complete surface wave band gap in piezoelectric phononic crystal," Phys. Rev. E **73**, 065601 (2006).
[7]T.-T. Wu, Z.-C. Hsu, and Z.-G. Huang, "Band gaps and the electromechanical coupling coefficient of a surface acoustic wave in a two-dimensional piezoelectric phononic crystal," Phys. Rev. B **71**, 064303 (2005).
[8]G. Monsivais, J. A. Otero, and H. Calas, "Surface and shear horizontal waves in piezoelectric composites," Phys. Rev. B **71**, 064101 (2005).
[9]I. V. Ostrovskii and A. B. Nadtochiy, "Domain resonance in two-dimensional periodically poled ferroelectric resonator," Appl. Phys. Lett. **86**, 222902 (2005).
[10]R. Lerch, "Simulation of piezoelectric devices by two- and three-dimensional finite elements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**(3), 233–247 (1990).

Predicting echo thresholds from speech onset characteristics

Scott D. Miller
Department of Psychology and Waisman Center, University of Wisconsin, Madison, Wisconsin 53706
sdmiller1@wisc.edu

Ruth Y. Litovsky
Department of Communicative Disorders and Waisman Center, University of Wisconsin, Madison, Wisconsin 53705
litovsky@waisman.wisc.edu

Keith R. Kluender
Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706
krkluend@wisc.edu

Abstract: Echo threshold variability has previously been examined using stimuli that are carefully controlled and artificial (e.g., clicks and noise bursts), while studies using speech stimuli have only reported average thresholds. To begin to understand how echo thresholds might vary among speech sounds, four syllables were selected in pairs that contrasted abruptness vs gradualness of onset envelopes. *Fusion* and *discrimination suppression thresholds*, two echo thresholds commonly used to study the *precedence effect*, differed among syllables. Results were used to evaluate two predictive heuristics adapted from *perceptual center (p-center)* models.

PACS numbers: 43.66.Pn, 43.71.Rt, 43.72.Dv, 43.66.Qp [QJF]
Date Received: November 23, 2008 **Date Accepted:** January 16, 2009

1. Introduction

Everyday listening conditions are characterized by multiple sound sources and numerous reflections. As a first step toward understanding auditory perception in reverberant conditions, studies have employed a simple experimental paradigm with only a single first-arriving or *lead* sound and a single later-arriving or *lag* sound. The lag is usually an exact replica of the lead, simulating the first reflection to reach the ear in a reverberant environment. At short lead-lag delays, the lead dominates or suppresses perception of lag information along a number of dimensions. This is known as the *precedence effect*. Examples of dimensions over which it operates include perception of the lag as a discrete event (usually assessed subjectively via a *fusion* task) and extraction of lag directional cues (assessed objectively via tasks such as *discrimination suppression*). One common goal has been to quantify echo thresholds, i.e., the lead-lag delay at which the lead ceases to dominate the lag.[1]

While some of the very first work in this area used continuous speech as one kind of stimulus,[2–4] most progress since then has been made using artificial stimuli, which allow greater experimental control but lack acoustic characteristics intrinsic to speech and other naturally produced sounds. The majority of precedence effect studies to date have used clicks or noise bursts that are only a few milliseconds in duration. Such stimuli do not result in temporal overlap between the lead and lag at inter-stimulus delays corresponding to psychophysical echo thresholds.[1] In contrast, a single syllable in speech is usually 150 ms or longer, much longer than echo thresholds reported in a small number of studies for complex sounds such as continuous speech and music.[3,5] Clicks and noise bursts have been generated in laboratories to eliminate periodicities and recognizable transients,[1] while pure tones have been used to eliminate spectral complexity.[6] Stimulus onset characteristics, which play a critical role in the precedence

effect,[6–8] are often held constant and have only been manipulated in a simplified linear fashion.[6] Information in speech is carried by just the kinds of spectral and temporal characteristics that are eliminated from most artificial stimuli.

Systematic variability among echo thresholds for various speech sounds has not previously been reported, and it is not clear how current auditory models might be best adapted to make accurate predictions about such differences. Precedence effect studies that have used speech stimuli[2,3,5] have done so with the implicit simplifying assumption of equivalent thresholds across various speech sounds, reporting only the average results across all speech stimuli used. This assumption has been explicitly built into the few precedence effect models that have attempted to account for experimental results with speech stimuli.[9,10] The current literature does not provide any greater precision in the characterization of speech echo thresholds than the following general observation from the first precedence effect report,[4] p. 335: "The interval over which fusion takes place is not the same for all kinds of sounds. The upper limit of the interval was found to be about 5 ms for single clicks and is apparently much longer, perhaps as much as 40 ms, for sounds of a complex character."

Existing literature points to some acoustic characteristics likely to be most relevant to variability among speech echo thresholds, should there prove to be any. In particular, binaural detection and localization results with non-speech stimuli have demonstrated the distinct importance of onset characteristics for both periodic (i.e., pure tone[6]) and aperiodic (i.e., noise[7,8]) stimuli. The other consistent finding has been the dominance of low-frequency information ($<$1.5 kHz) over high-frequency information in precedence effect phenomena.[11,12]

One can also look to the speech perception literature, where waveform onsets and low-frequency information have emerged as critical predictors of so-called *perceptual centers* or *p-centers*.[13] The p-center is the beat or perceived moment of occurrence of a sound event as distinguished from the absolute acoustic onset.[14] From the p-center literature, we adopted Scott's[13] *frequency-dependent amplitude increase model* (FAIM) for the prediction of speech echo thresholds. Specifically, we replicated Scott's[13] procedure for the calculation of *rise time* values, the metric determined to most accurately predict p-centers. The extension of rise times to the prediction of echo thresholds was consistent with Rakerd and Hartmann's[6] identification of *onset rate* (increase in sound pressure per unit time) as the best predictor of binaural localization ability, but a frequency weighted application of this principle seemed more promising for use with naturally produced sounds. In addition, FAIM rise times have the advantages of being clearly specified and relatively easy to implement with speech stimuli.

The current study addressed two separate but related questions: (1) do echo thresholds for speech sounds vary systematically, and (2) can we predict these thresholds from acoustic characteristics? The first was addressed by applying a repeated-measures analysis of variance (ANOVA) to echo thresholds for different speech sounds; the second required a predictive heuristic, in this case FAIM rise times, and was addressed via correlational analysis.

2. Method

2.1 Participants

Nine young adults, six females and three males, initially volunteered to participate. To continue in the study, participants were required to pass a 20 dB hearing screening at 500, 1000, 2000, 4000, and 8000 Hz for each ear and have a symmetrical hearing profile, defined as no more than a 10 dB discrepancy between right and left ear thresholds at a particular frequency. None had any history of hearing difficulties, but one male was deemed ineligible for participation due to an asymmetrical hearing profile. One female did not return for testing after the initial practice session, leaving a final group of seven participants. All participants spoke English as their first and primary language.

2.2 Stimuli

Four syllables were recorded from an adult female: [bʌ], [dʌ], [wʌ], and [yʌ] (transliterated: "buh," "duh," "wuh," and "yuh"). These stimuli were selected as pairs, [bʌ] vs [wʌ] and [dʌ] vs

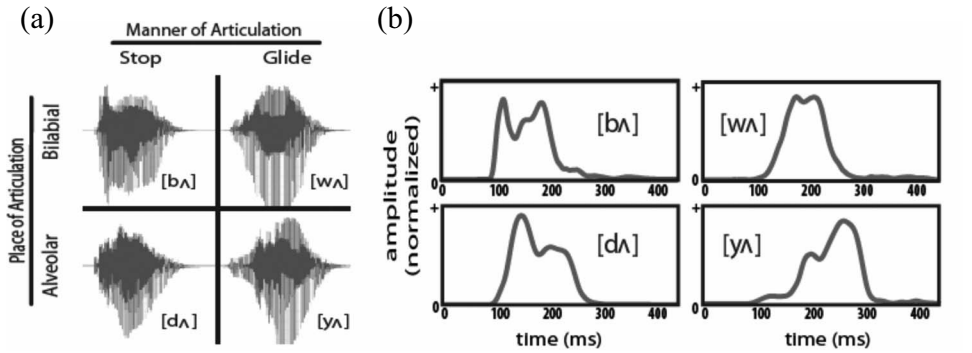


Fig. 1. (a) Raw waveforms of four syllables used as stimuli. Stimuli were digitally sampled at 44.1 kHz, and amplitude levels were root-mean-square equalized (see Multimedia link in text). (b) Outputs from filtering procedure used to calculate onset rise times for each stimulus.

[yʌ], which would differ significantly in terms of rise times but would otherwise have similar spectral and temporal characteristics. The voiced stop consonants in [bʌ] and [dʌ] have abrupt waveform onsets compared to the glides in [wʌ] and [yʌ], which have gradual onsets [see Fig. 1(a)].

[Mm. 1. [bʌ], [wʌ], [dʌ], and [yʌ] stimulus files concatenated in wav format (134 KB).]

Rise times were calculated following the procedure of Scott.[13] Each syllable was passed through a gammatone filter[15] with a center frequency of 578 Hz and an equivalent rectangular bandwidth of 4.0. These outputs, which contained significant information several hundred hertz above and below the center frequency, were full-wave rectified and smoothed with a 25 Hz Butterworth filter [see Fig. 1(b) for results]. The onset time (in milliseconds) between 10% and 90% of the maximum amplitude of each smoothed waveform was calculated, producing a frequency-dependent rise time for each recorded syllable. Rise times (in parentheses following each syllable) predicted that echo thresholds would be ordered as follows, from lowest to highest: [bʌ] (14.83), [dʌ] (33.22), [wʌ] (43.65), and [yʌ] (91.54).

2.3 Equipment

A customized MATLAB program was developed for randomization and presentation of stimuli, which was accomplished via a Tucker-Davis Technologies System 3 multiple input-output processor. Testing took place in a standard IAC sound booth (reverberation time, T_{60}=250 ms), where an array of loudspeakers (Cambridge SoundWorks Henry Kloss Center/Surround IV) was positioned in the horizontal plane. Each speaker in the semi-circular array was at ear level, 1.4 m from the approximate center of the seated participant's head. A computer monitor was placed under the front loudspeaker, and participants entered responses via a mouse by clicking on appropriate icons as instructed.

2.4 Design and procedure

Two tasks were used to measure echo thresholds: fusion and discrimination suppression. All seven participants completed both tasks. Before data collection began, each participant completed a 2-h practice session to establish a consistent response criterion on each task. The procedure for each task and thresholds (50% two sounds perceived on the fusion task, and 70.9% correct, i.e., $d'=1.0$, on the discrimination task) was similar to those described by Freyman *et al.*[16] The following delays, which always produced temporal overlap between lead and lag, were used for both tasks: 1, 5, 10, 15, 20, 25, 30, 35, 40, and 45 ms. For both tasks the following conditions applied: (a) The same stimulus was always presented as both lead and lag at 60 dBA, (b) participants were instructed to face the computer monitor in front (0° azimuth), (c) the lead was always from −45° azimuth, and (d) participants initiated each trial at their own pace using

Table 1. Both raw and centered group-mean thresholds are reported for each syllable on each task. Centered values are relative to the task mean, facilitating cross-task comparisons. Syllable error terms are *within subjects standard deviations* (relative to centered thresholds), consistent with the design of the study. Task mean error terms are *between subjects standard deviations* (relative to raw thresholds).

	Task									
	Fusion					Discrimination				
Group mean	[bʌ]	[dʌ]	[wʌ]	[yʌ]	Task mean	[bʌ]	[dʌ]	[wʌ]	[yʌ]	Task mean
Raw	17.00	14.79	19.23	21.73	18.19	10.91	6.85	13.26	13.83	11.14
Centered	−1.19	−3.40	1.04	3.54	18.19	−0.23	−4.29	2.12	2.69	11.14
Std. dev.	2.36	3.30	2.35	2.15	6.79	3.14	3.33	4.53	2.53	5.72

onscreen controls. During the fusion task, the lag was always presented from +45° azimuth. Participants, who were informed that the task involved an auditory "illusion" and that two sounds were always presented, made a subjective decision as to whether they perceived there to be one or two sounds. No feedback could be given, but participants were instructed to only select two if they were certain they had heard a second sound coming from a location to the right of center. During the discrimination task, the lag was presented from either +35° or +55°. Participants were instructed to decide whether the lag originated from the left or right of a visual reference point at +45° (a two-alternative forced choice), and feedback was given after each response.

The method of constant stimuli was used, and blocks alternated between tasks in a counterbalanced manner. Within each block, all stimuli and delays were used an equal number of times, and the presentation order was completely randomized across stimuli and delays; lag locations were included in this block randomization on the discrimination task. Across all blocks of the fusion task, 20 repetitions were completed for each stimulus at each delay, resulting in 800 total trials. For the discrimination task, 40 repetitions (20 from each lag location) resulted in a total of 1600 trials. Participants were allowed to take breaks as often and for as long as they wanted during testing. They were encouraged to take a minimum of one break every 200 trials. Trials were typically completed across three 2-h testing sessions spaced over several days.

3. Results and discussion

3.1 Echo thresholds

Group mean thresholds are reported in Table 1. Task means, 18.19 ms for fusion and 11.14 ms for discrimination suppression, were slightly higher than those reported in the literature for click stimuli (e.g., 5–9 ms for thresholds on both tasks[16] using identical threshold criteria and free field procedures) but slightly lower than fusion thresholds typically reported for connected speech (e.g., 30–50 ms[3]). Litovsky *et al.*[1] provided a review of thresholds found on both tasks with different types of stimuli. In the present study, the use of single syllables as stimuli, rather than sentences or concatenated word lists, probably explains why thresholds fell somewhere midway between previous results with artificial stimuli, on the one hand, and continuous speech, on the other.

3.2 Analysis of variance

For all statistical tests, a level of $p < 0.05$ was used to determine significance. When task (fusion vs discrimination), manner of articulation (stop vs glide), and place of articulation (bilabial vs alveolar) were entered as factors in a three-way repeated-measures ANOVA, there was a significant main effect of manner, $F(1,6) = 11.45$, $p = 0.02$, and a significant manner by place interac-

tion, $F(1,6)=15.67$, $p=0.01$. No other main effects or interactions were significant. This supported our primary hypothesis that echo thresholds for speech sounds would differ across syllables.

The main effect of manner was in the predicted direction, with stops as a class having lower group mean thresholds than glides, while the interaction of place with manner affected the strength but not the direction of this effect. Therefore, this analysis of onset effects, at the admittedly gross level of syllable category, also provided support for our second hypothesis that the abrupt onsets of stops would tend to result in lower thresholds. Paired-sample *t*-tests reached significance for [dʌ] vs [yʌ], [bʌ] vs [yʌ], and [wʌ] vs [yʌ] on the fusion task and [dʌ] vs [bʌ] and [dʌ] vs [yʌ] on the discrimination task.

3.3 Correlational analyses

Correlations between rise times and group mean echo thresholds were in the predicted direction, but neither slope was statistically significant: $r=0.82$, $r^2=0.67$, and $p=0.18$ for the fusion task, and $r=0.58$, $r^2=0.33$, and $p=0.42$ for the discrimination task. Although the small number of data points meant statistical power was low for this analysis, rise times have two characteristics that may account for their failure to predict observed differences among echo thresholds more accurately. They falsely assume linearity in waveform onsets, and they do not relate onset characteristics to the waveform as a whole.[17] An alternative p-center metric, Howell's[18] *center of gravity* can be adapted to address both of these potential shortcomings. It had never been applied in a frequency-specific manner, but the limited success of frequency-dependent rise times in the current study seemed to warrant the test of a hybrid *frequency-dependent center of gravity* (FCoG) model. In deriving FCoGs, Scott's[13] filtering procedure was again used—and consistent with this procedure, the syllable was defined as beginning and ending at the points where the waveform amplitude was equal to 10% of the maximum amplitude—but the remaining steps were suggested by Howell.[19] The duration from the beginning of the syllable to the point at which the integral (area under the curve) equaled one-half of the integral for the entire syllable was expressed as a ratio over the duration of the entire syllable, i.e., as a unitless proportion. From a signal detection theoretical perspective, this would seem to capture the detectability of the lag onset as well as the masking potential of the continuous and offset portions of the lead in a single metric.

The FCoGs (in parentheses following each syllable) predicted that echo thresholds would be ordered as follows, from lowest to highest: [dʌ] (0.22), [bʌ] (0.41), [wʌ] (0.49), and [yʌ] (0.59). Correlations for the fusion task, $r=0.976$, $r^2=0.952$, and $p=0.02$, and the discrimination task, $r=0.982$, $r^2=0.970$, and $p=0.02$, indicated that FCoGs accounted for almost all of the variance in echo thresholds. Both slopes were significant even after a conservative Bonferroni correction was applied to this *post hoc* analysis.

4. Conclusions

In this investigation, echo thresholds differed as a function of syllable-initial phonetic features and correlated significantly with FCoGs, shedding light on properties of speech stimuli worth exploring further in relation to the precedence effect and speech perception in reverberant situations. Predictive heuristics tested in this study fit most naturally with other signal detection theoretical approaches in the binaural literature (see Saberi and Petrosyan[20] for review) and in the broader perceptual sciences in that they focus on the information-bearing properties of the stimulus. This approach is less popular than physiologically oriented modeling in the binaural literature (see Stern and Trahiotis[21] for review), which focuses on the transformational processes accomplished by the auditory system, but it has shown merit here as an entrée into the characterization of variability among echo thresholds for speech stimuli. In the long run, this complementary approach should facilitate improvements in physiological models as well.

Without experimental data from a much wider variety of stimuli (e.g., different phoneme combinations and talkers), it is difficult to know whether FCoGs or FAIM rise times can be used to predict all speech echo thresholds. Perhaps the most critical test for any predictor will come with initial voiceless consonants, in particular voiceless fricatives. These onsets present a

challenge because they lack both the low-frequency spectral peaks and periodicities characteristic of voiced consonant and vowel sounds. There are no explicit models for how high-frequency-aperiodic and low-frequency-periodic cues are combined to produce precedence effect phenomena when both are sequentially present in a single stimulus, as they are, for example, in syllables that begin with voiceless fricatives. These challenges may require a more complex model, such as the p-center model of Pampino–Marschall,[22] which determines *partial onset events* as well as *syllabic centers of gravity* from frequency-specific loudness dynamics within multiple critical bands. If multivariate predictors prove to be necessary, their application to continuous stimuli may benefit from a learning algorithm and training procedure (e.g., the neural network model of Wilson and Darrell[23]).

Acknowledgments

This work was supported by grants from the NIH-NIDCD (Grant No. R01 DC030083 to R.Y.L. and T32 DC005359 to Susan Ellis Weismer, University of Wisconsin, Madison).

References and links

[1]R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654 (1999).
[2]H. Haas, "The influence of a single echo on the audibility of speech," J. Audio Eng. Soc. **20**, 145–159 (1972).
[3]J. P. A. Lochner and J. F. Burger, "The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of speech," Acustica **8**, 1–10 (1958).
[4]H. Wallach, E. B. Newman, and M. R. Rosenzweig, "The precedence effect in sound localization," Am. J. Psychol. **62**, 315–336 (1949).
[5]B. Rakerd, W. M. Hartmann, and J. Hsu, "Echo suppression in the horizontal and median sagittal planes," J. Acoust. Soc. Am. **107**, 1061–1064 (2000).
[6]B. Rakerd and W. M. Hartmann, "Localization of sound in rooms, III: Onset and duration effects," J. Acoust. Soc. Am. **80**, 1695–1706 (1986).
[7]T. Houtgast and R. Plomp, "Lateralization threshold of a signal in noise," J. Acoust. Soc. Am. **44**, 807–812 (1968).
[8]T. Houtgast and S. Aoki, "Stimulus-onset dominance in the perception of binaural information," Hear. Res. **72**, 29–36 (1994).
[9]O. Schwartz, J. G. Harris, and J. C. Principe, "Modeling the precedence effect for speech using the gamma filter," Neural Networks **12**, 409–417 (1999).
[10]B. Supper, T. Brookes, and F. Rumsey, "A new approach to detecting auditory onsets within a binaural stream," Paper presented at the 114th Convention of the Audio Eng. Soc., Amsterdam, Netherlands (March 2003). Retrieved from http://epubs.surrey.ac.uk/recording/23 (Last viewed June 25, 2008).
[11]B. G. Shinn-Cunningham, P. M. Zurek, N. I. Durlach, and R. K. Clifton, "Cross frequency interactions in the precedence effect," J. Acoust. Soc. Am. **98**, 164–171 (1995).
[12]D. J. Tollin and G. B. Henning, "Some aspects of the lateralization of echoed sound in man. II. The role of the stimulus spectrum," J. Acoust. Soc. Am. **105**, 838–849 (1999).
[13]S. K. Scott, "P-centers in speech: An acoustic analysis," Ph.D. dissertation, University College London, UK, 1993.
[14]J. Morton, S. Marcus, and C. Frankish, "Perceptual centers (p-centers)," Psychol. Rev. **83**, 405–408 (1976).
[15]M. Slaney, "An efficient implementation of the Patterson-Holdsworth auditory filterbank," Apple Computer Technical Report No. 35; Retrieved from http://citeseer.ist.psu.edu/8863.html (Last viewed June 25, 2008).
[16]R. L. Freyman, R. K. Clifton, and R. Y. Litovsky, "Dynamic processes in the precedence effect," J. Acoust. Soc. Am. **90**, 874–884 (1991).
[17]While rise times accurately predict p-centers of natural stimuli (Ref. 13), perception of onsets is complicated by the presence of background sounds (Ref. 24), and it remains an open question whether absolute onsets are better than relative onsets for predicting echo thresholds of natural stimuli in reverberant conditions.
[18]P. Howell, "Prediction of the p-center from the distribution of energy in the amplitude envelope: I," Percept. Psychophys. **43**, 90–93 (1988).
[19]Howell himself (Ref. 18) used a procedure to calculate centers of gravity that assumed linearity in the amplitude envelope by modeling the syllable onset as a triangle and the remainder of the syllable as a rectangle, but he suggested an alternative procedure that takes integrals directly from rectified complex waveforms and does not assume linearity. This alternative is more straightforward theoretically, but the computations are more complex.
[20]K. Saberi and A. Petrosyan, "A detection-theoretic model of echo inhibition," Psychol. Rev. **111**, 52–66 (2004).
[21]R. M. Stern and C. Trahiotis, "Models of binaural interaction," in *Handbook of Perception and Cognition: Vol. 4. Hearing*, 2nd ed., edited by B. C. J. Moore (Academic, New York, 1995), pp. 347–386.

[22]B. Pompino-Marschall, "On the psychoacoustic nature of the P-center phenomenon," J. Phonetics **17**, 175–192 (1989).
[23]K. W. Wilson and T. Darrell, "Learning a precedence effect-like weighting function for the generalized cross-correlation framework," IEEE Trans. Audio, Speech, Lang. Process. **14**, 2156–2164 (2006).
[24]J. Vos and R. Rasch, "The perceptual onset of musical tones," Percept. Psychophys. **29**, 323–335 (1981).

Vocal tract modes based on multiple area function sets from one speaker

Brad H. Story
Speech Acoustics Laboratory, Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona 85721
bstory@u.arizona.edu

Abstract: The purpose of this study was to derive vocal tract modes from a wider range of vowel area functions for a specific speaker than has been previously reported. Area functions from Story *et al.* [(1996). J. Acoust. Soc. Am. **100**, 537–554] and Story [(2008). J. Acoust. Soc. Am. **123**, 327–335] were combined in a composite set from which modes were derived with principal component analysis. Along with scaling coefficients, these modes were used to generate a [F1, F2] formant space. In comparison to formant spaces similarly generated based on the two area function sets alone, the combined version provides a wider range of both F1 and F2 values. This new set of modes may be useful for inverse mapping of formant frequencies to area functions or for modeling of vocal tract shape changes.

PACS numbers: 43.70.Bk, 43.70.Aj [AL]
Date Received: September 7, 2008 **Date Accepted:** January 8, 2009

1. Introduction

For production of vowels, the vocal tract area function has been shown to be fairly well represented by two canonical deformation patterns, or "modes," (e.g., Story and Titze, 1998; Story, 2005b; Mokhtari *et al.*, 2007) derived from principal component analysis (PCA). Such modes have been derived from speaker-specific area function sets, but are similarly shaped (in terms of their variation along the length of the vocal tract) across speakers, and are related to specific patterns of formant frequencies when appropriately scaled and superimposed on a mean area function.

The modal representation of the vocal tract shape allows for an essentially one-to-one mapping between the scaling coefficients of the modes and the first two formant frequencies and has led to a method for mapping time-varying formant frequencies extracted from recorded speech to a time-varying sequence of area functions (Story and Titze, 1998; Story and Titze, 2002; Mokhtari *et al.*, 2007). Development of a kinematic model of the vocal tract area function that can be used to simulate speech is also based on this same modal representation (Story, 2005a). Although both the inverse mapping technique and kinematic model have been shown to be reasonably successful at bridging the area function-to-acoustic relation, the ability of the modal representation to produce a wide range of vocal tract shapes is limited by the original set of area functions on which it is based. For example, the modal representation derived in Story and Titze (1998) and subsequently utilized in Story and Titze (2002) and Story (2004, 2005b) was based on the ten vowel area functions reported for an adult male (Story *et al.*, 1996). Hence, the boundaries of the potential [F1, F2] vowel space afforded by the modal representation are essentially defined by the formant frequencies of those original area functions. Although additional sets of speaker-specific area functions reported by Story (2005b) and Mokhtari *et al.* (2007) have been used to derive modal representations of individual speakers, they are also limited by the range of formant frequencies produced by the original area functions obtained from each speaker. This is not necessarily a severe limitation assuming that each speaker produced hyperarticulated vowels (thus, producing extreme [F1, F2] formant values). If the vowels tend toward being centralized, however, the working space of a modal-based vocal tract model will be constrained.

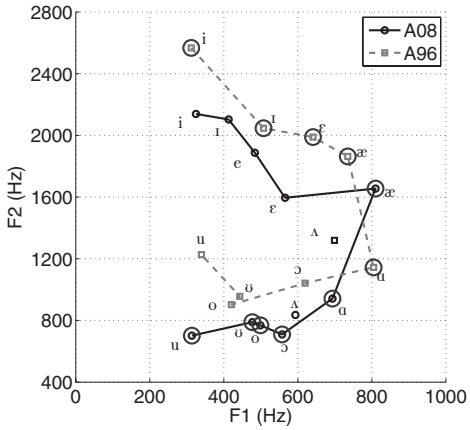


Fig. 1. Vowel space plot of F1 and F2 frequencies calculated from the A96 and A08 versions of each vowel (see text). The data points are connected by solid or dashed lines to clarify the set to which they belong and to provide a rough outline of the possible vowel space. Because of their positions in the F2 vs F1 plane, the two [ʌ] vowels are not connected to the other vowels within their respective sets. The large open circles denote those vowels chosen from each set that were combined in a composite set.

Recently, Story (2008) reported an additional set of 11 vowel area functions for the same speaker who produced the 10 vowels originally reported in Story *et al.* (1996). F1 and F2 formant values were determined from the frequency response function calculated for each area function in the two sets. It was noted that, in general, the F2 values calculated for the new area function set were shifted downward in frequency relative to those calculated for the area functions of,[1] Story *et al.* (1996) as shown in Fig. 1. It was suggested that this difference originated from a tendency, in the new area functions, to slightly constrict the pharyngeal portion of the vocal tract while expanding the oral cavity. Since the new area functions are configured somewhat differently than the 1996 versions and produce different [F1, F2] values, they provide additional samples of the same speaker's vocal tract that might be *combined* with the original set in a PCA. The resulting vocal tract modes and mean area function would then be representative of a wider range of vowel configurations than either area function set alone.

The purpose of this letter is to demonstrate that a modal representation derived from a particular combination of area functions selected from the data of Story *et al.* (1996) and Story (2008) produces an [F1, F2] vowel space that is larger than that produced by modes derived from either area function set alone.

2. Principal component analysis

Area function sets from both Story *et al.* (1996) and Story (2008) were used in the analysis. The area functions of Story *et al.* (1996) included the vowels [i ɪ ɛ æ ʌ ɑ ɔ o ʊ u] and will henceforth be referred to as "A96." In addition to these same ten vowels, the Story (2008) data include the vowel [e]; this set of 11 area functions will be referred to as "A08."

Any given area *function* can be represented by two components, an area vector and a length increment. The area *vector* $A(i)$ contains 44 cross-sectional areas,[2] assumed to represent a concatenation of tubelets ordered consecutively from glottis to lips. The index i denotes this ordering and extends from 1 to 44. The length increment Δ is the distance between consecutive cross-sectional areas and can be considered to be the tubelet length. The 44-section area vectors and length increments, as used in the present study, were reported in Story and Titze (1998) for A96[3] and in Story (2008) for A08. The A96 set, however, was smoothed for this study in the same manner as the A08 set (see Story, 2008, p. 328).

As mentioned in the Introduction, the calculated [F1, F2] formant frequencies for each of the two area function sets are shown in Fig. 1. The open circles indicate the vowels in the plot

that are representative of the boundaries of the displayed vowel space, regardless of set membership, and their corresponding area functions served as the combined data set on which a new PCA was performed. This combined set is referred to as "A9608x" and specifically includes [i ɪ ɛ æ ɑ] from A96 and [æ ɑ ɔ o ʊ u] from A08, where the "x" denotes that the combined set excludes some of the vowels in A96 and A08. Note that the target vowels [æ] and [ɑ] from both A96 and A08 are included in the combined set because they represent part of the overall boundary of the vowel space.

The collection of area vectors for a given set (i.e., A96, A08, or A9608x) can be represented in matrix form as $A(i,j)$, where i is the area index and j denotes the particular vowel in the clockwise order indicated by the points along the dashed or solid lines for A96 and A08, respectively, or by the large circles in Fig. 1 for A9608x (i.e., $j=1$ for [i], $j=2$ for [ɪ],…, $j=11$ for [u]). Following Story (2005b), the PCA was performed on the equivalent diameters of the cross-sectional areas rather than on the areas themselves. In addition, the *length* of each area function was included in the same manner as Yehia *et al.* (1996) and Mokhtari *et al.* (2007) where the variance of the length increment Δ_j is first normalized by the largest variance of the equivalent diameters, and then becomes the 45th element. Thus, a matrix $D(i,j)$, containing diameter and length information, is constructed as

$$D(i,j)=\begin{cases}\sqrt{\dfrac{4}{\pi}A(i,j)} & \text{for } i=1,\ldots,44\\[2ex] \dfrac{(\Delta_j-\bar{\Delta})\sigma_D}{\sigma_\Delta}+\bar{\Delta} & \text{for } i=45,\end{cases} \tag{1}$$

where Δ_j is the length increment of area vector j, $\bar{\Delta}$ is the mean length increment, σ_Δ is the standard deviation of the length increments, and σ_D is the largest standard deviation within any section i of the equivalent diameters.

Matrix $D(i,j)$ can be represented by a mean and variable part,

$$D(i,j)=\Omega(i)+\alpha(i,j), \tag{2}$$

where $\Omega(i)$ is the mean vector across $D(i,j)$, and $\alpha(i,j)$ is the variation superimposed on $\Omega(i)$ to produce a specific vector. The PCA was then carried out by calculating the eigenvectors of a covariance matrix formed with $\alpha(i,j)$. The specific implementation was essentially the same method as reported in Story and Titze (1998), and results in the following parametric representation of the original $D(i,j)$ matrix:

$$\hat{D}(i,j)=\left[\Omega(i)+\sum_{i=1}^{N}q_i(j)\phi_i(i)\right],\quad i=[1,N]\quad (N=45),\quad j=[1,11], \tag{3}$$

where the $\phi_i(i)$'s are 45-element eigenvectors (modes) that, when multiplied by the appropriate scaling coefficients $q_i(j)$, will reconstruct each area vector in $A(i,j)$ for $i=[1,44]$ and each length increment for $i=45$ by the following:

$$\hat{A}(i,j)=\frac{\pi}{4}\hat{D}(i,j)^2,\quad i=[1,44],\quad j=[1,11], \tag{4a}$$

$$\hat{\Delta}(j)=\frac{(\hat{D}(45,j)-\bar{\Delta})\sigma_\Delta}{\sigma_D}+\bar{\Delta},\quad j=[1,11]. \tag{4b}$$

Area functions generated from the mean diameter function $\Omega(i)$ [i.e., $q_i(j)=0$] of each set A96, A08, and A9608x are plotted against the normalized distance[4] from the glottis in Fig. 2(a). It is noted that the area function based on the combined set (thick line) is more constricted in the

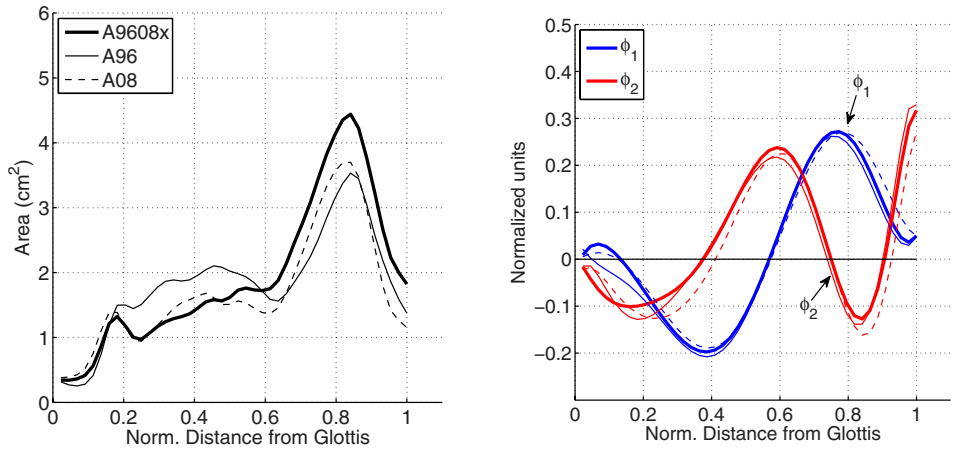


Fig. 2. (Color online) Mean area functions and modes derived for each of the three area function sets A96, A08, and A9608x: (a) mean area functions and (b) ϕ_1 and ϕ_2 modes where the thick line corresponds to set A9608x, the thin line to A96, and the dashed line to A08.

pharyngeal portion (from approximately 0.2 to 0.6 along the length) and more expanded in the oral portion (from 0.6 to 1) than either of the other two.

The two modes that accounted for most of the variance in the analysis of the three area function sets [referred to as $\phi_1(i)$ and $\phi_2(i)$] are shown in Fig. 2(b), again plotted against the normalized distance from the glottis. Note that the modes have been smoothed by fitting them with eighth-order polynomials. This simplifies the visual comparison of the modes but maintains their gross characteristics. As in Fig. 2(a), the thick, thin, and dashed lines represent the A9608x, A96, and A08 sets, respectively. The variances accounted for by each mode in each of the three sets are given in Table 1. The first mode ϕ_1 accounts for greater than 64% of the variance, whereas ϕ_2 accounts for about 20%. The total amount of the variance accounted for by these two modes was greatest for the A9608x set with 89.6%. Qualitatively there are some minor differences in the amplitude of the modes across the three sets, but the overall shape is essentially the same. That is, with a positive scaling coefficient, ϕ_1 will expand the front half of the vocal tract and constrict the back half, and vice-versa with a negative scaling coefficient. When positively scaled, the second mode imposes an expansion near the lips, followed by a constriction, an expansion, and another constriction above the glottis; the opposite effect occurs with a negative scaling coefficient.

3. Vowel space calculations

A mapping was generated for each area function set that relates the two scaling coefficients q_1 and q_2 to the first two formant frequencies (F1 and F2). Based on the PCA performed on each

Table 1. Percentage of the total variance in the matrix $D(i,j)$ accounted for by the two most significant modes (in their smoothed form) calculated for each of the sets A96, A08, and A9608x.

Mode	A96	A08	A9608x
Φ_1	64.8	69.8	65.9
Φ_2	21.3	18.5	23.7
Total	86.1	88.3	89.6

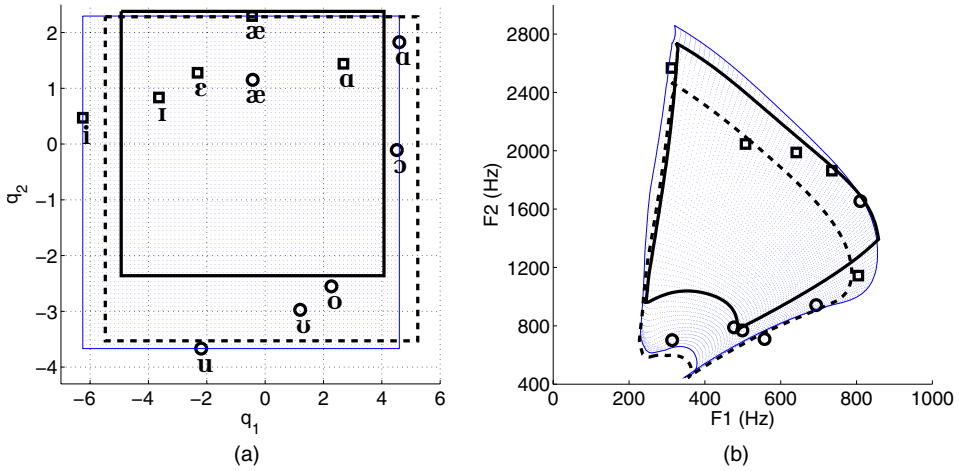


Fig. 3. (Color online) Coefficient and [F1, F2] grids. (a) The grid in the background, bounded by the thin line, represents the coefficient pairs generated for the A9608x set. The squares denote the coefficient pairs corresponding to the vowels chosen from the A96 set whereas the circles are those chosen from the A08 set. The thick solid and dashed lines indicate the outlines of the coefficient grids generated separately for the A96 (solid) and A08 (dashed) sets, respectively. (b) The deformed grid in the background, bounded by the thin line, represents the [F1, F2] space generated from the coefficient grid in (a). The squares and circles denote the calculated [F1, F2] values for the A96 and A08 vowels, respectively. The thick solid and dashed lines represent the outlines of the [F1, F2] spaces generated from coefficient grids for the A96 and A08 sets, respectively.

set, an equal increment continuum was generated for each mode coefficient that ranged from their respective minimum to maximum values. The increments were specified as

$$\Delta q_1 = \frac{q_1^{\max} - q_1^{\min}}{M-1}, \tag{5a}$$

$$\Delta q_2 = \frac{q_2^{\max} - q_2^{\min}}{N-1}, \tag{5b}$$

where M and N are the numbers of increments along each coefficient dimension. For this study $M=N=80$. The coefficient continua were then generated by

$$q_{1m} = q_1^{\min} + m\Delta q_1, \quad m = 0, \ldots, M-1, \tag{6a}$$

$$q_{2n} = q_2^{\min} + n\Delta q_2, \quad n = 0, \ldots, N-1. \tag{6b}$$

with m and n serving as indices along each continuum. By modifying Eq. (3) to contain only two modes and eliminating the dependence on a specific vowel (i.e., jth vowel), an area vector and length increment can be directly generated with

$$A_{mn}(i) = \frac{\pi}{4}[\Omega(i) + q_{1m}\phi_1(i) + q_{2n}\phi_2(i)]^2, \quad i = [1,44], \tag{7a}$$

$$\Delta_{mn} = \frac{(\Omega(i) + q_{1m}\phi_1(i) + q_{2n}\phi_2(i) - \bar{\Delta})\sigma_\Delta}{\sigma_D} + \bar{\Delta}, \quad i = 45. \tag{7b}$$

Shown in Fig. 3(a) is the 80×80 coefficient grid for the A9608x set along with the coefficient pairs that reconstruct the original 11 vowels used in the PCA for this set. The squares and circles represent those vowels extracted from the A96 and A08 sets, respectively, to make up the combined set. Also shown in this figure are the *outlines* of coefficient grids based on the separate

PCAs of the A96 (solid line) and A08 (dashed line) sets. The coefficient ranges are similar for the A9608x and A08 sets, but the range for the A96 set is contracted relative to the other two. It is noted that all three sets have nearly the same upper limit of the q_2 coefficient.

For each of the 6400 coefficient pairs in the grid (i.e., every intersection point), an area vector and length increment were generated. A frequency response function of each resulting area function was then calculated with a frequency-domain technique based on cascaded "ABCD" matrices (Sondhi and Schroeter, 1987; Story *et al.*, 2000). This calculation included energy losses due to yielding walls, viscosity, heat conduction, and acoustic radiation at the lips; side branches such as the piriform sinuses were not considered. Formant frequencies (F1 and F2) were determined by finding the peaks in the frequency response functions with an automated peak-picking algorithm (Titze *et al.*, 1987). These operations were carried out separately for the coefficient grids corresponding to all three area function sets (A96, A08, and A9608x).

The resulting formant space for A9608x is shown in Fig. 3(b). With the exception of a small amount of overlap along the upper edge, this represents a one-to-one mapping between $[q_1, q_2]$ coefficient pairs and [F1, F2] formant pairs, similar to those demonstrated in previous publications (e.g., Story and Titze, 1998; Story, 2005b). The formant pairs corresponding to the 11 vowel area functions in the A9608x set are shown as squares and circles. The formant grid outlines for the A96 and A08 sets are also shown in this figure as solid and dashed lines, respectively. Note that the A08 outline is shifted downward relative to that for A96, as would be expected based on the data shown previously in Fig. 1. More importantly, the formant grid based on A9608x encompasses nearly all of the space outlined by both the A96 and A08 sets.

4. Discussion

Finding the formant grid for the A9608x to be larger than that calculated for either the A96 or A08 sets suggests that the modes based on this combined set may be more useful for purposes of inverse mapping and vocal tract modeling. In the case of inverse mapping, a wider variation in [F1, F2] formant frequencies can be accommodated within the formant space and thus related to a wider range of vocal tract area function shapes. For modeling vocal tract shape change as a component of synthesizing speech (cf. Story, 2005a), the modes from the combined set allow for production of a wider range of formant frequencies than those from the other two sets, and could potentially facilitate more natural sounding synthesis.

The similarity of the mode shapes across each of the three area function sets (A96, A08, and A9608x) does, however, raise the question of whether it would be possible to produce an expanded formant space for either the A96 or A08 sets simply by allowing the mode coefficient values to exceed their ranges produced by the PCA? For example, could the A96 modes and mean diameter function produce a formant space similar to the grid shown in Fig. 3 if the A9608x coefficient space was used instead of the more limited set of A96 coefficients? It is almost certain that expanding the coefficient ranges will have the effect of enlarging the formant space. In fact, Story (2004) used this approach to model a compensation for a lip-tube constraint imposed on the vocal tract shape. But expanding the coefficient range will always be limited by the degree of constriction produced within the vocal tract. That is, increasing the magnitude of either the q_1 or q_2 coefficients will ultimately result in occlusion of the vocal tract, and possibly generating unrealistically short or long tract lengths.

Acknowledgments

This research was supported by NIH Grant No. R01-DC04789.

References and links

[1]F1 and F2 formant frequencies measured from audio recordings also indicated the same downward shift of F2 for the Story (2008) study relative to Story *et al.* (1996). The magnitude of the shift, however, was less than that observed in the calculated formants.

[2]The use of 44 elements to represent an area function has been adopted by the author in previous publications, but is not a requirement. The number derives from the approximate spatial resolution obtained in MRI-based recon-

structions of vocal tract shape (e.g., Story *et al.*, 1996). It is also convenient to use 44 elements for simulating speech with acoustic waveguide models because it allows for a sampling frequency of 44.1 kHz when the tract length is approximately 17.5 cm (typical adult male).

[3]The area functions in the A96 set were indeed originally reported in Story *et al.* (1996) but were provided in a form with a variable number of sections (tubelets) where the length increment was constant across all vowels. Story and Titze (1998) resampled the area functions so that each area vector consisted of 44 sections, but the length increment was vowel dependent.

[4]Normalized distance is used for displaying the reconstructed area functions and modes because the actual vocal tract length is variable and depends on a given combination of mode scaling coefficients [see Eq. (7)].

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.

Sondhi, M. M., and Schroeter, J. (**1987**). "A hybrid time-frequency domain articulatory speech synthesizer," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 955–967.

Story, B. H. (**2004**). "On the ability of a physiologically-constrained area function model of the vocal tract to produce normal formant patterns under perturbed conditions," J. Acoust. Soc. Am. **115**, 1760–1770.

Story, B. H. (**2005a**). "A parametric model of the vocal tract area function for vowel and consonant simulation," J. Acoust. Soc. Am. **117**, 3231–3254.

Story, B. H. (**2005b**). "Synergistic modes of vocal tract articulation for American English vowels," J. Acoust. Soc. Am. **118**, 3834–3859.

Story, B. H. (**2008**). "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002," J. Acoust. Soc. Am. **123**, 327–335.

Story, B. H., Laukkanen, A.-M., and Titze, I. R. (**2000**). "Acoustic impedance of an artificially lengthened and constricted vocal tract," J. Voice **14**, 455–469.

Story, B. H., and Titze, I. R. (**1998**). "Parameterization of vocal tract area functions by empirical orthogonal modes," J. Phonetics **26**, 223–260.

Story, B. H., and Titze, I. R. (**2002**). "A preliminary study of voice quality transformation based on modifications to the neutral vocal tract area function," J. Phonetics **30**, 485–509.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**, 537–554.

Titze, I. R., Horii, Y., and Scherer, R. C. (**1987**). "Some technical considerations in voice perturbation measurements," J. Speech Hear. Res. **30**, 252–260.

Yehia, H. C., Takeda, K., and Itakura, F. (**1996**). "An acoustically oriented vocal-tract model," IEICE Trans. Inf. Syst. **E79-D**, 1198–1208.

A portable infrasound generator

Joseph Park and James Robertson
Infrasound Laboratory, University of Hawaii, 73-4460 Queen Kaahumanu Highway No. 119, Kailua-Kona, Hawaii 96740
jpark@isla.hawaii.edu, james@isla.hawaii.edu

Abstract: The rotary subwoofer is a novel low frequency transducer capable of efficiently generating infrasound from a compact source. A field-deployable version of this device may find application as a calibration source for infrasound arrays of the International Monitoring System (IMS) [(2001). *The Global Verification Regime and the International Monitoring System* (CTBTO Preparatory Commission Vienna International Centre, Vienna, Austria)]. A prototype tested at the IMS infrasound array I59US demonstrated the ability to insonify all elements of the array from a standoff distance of 3.8 km. Signal-to-noise ratios of continuous wave signals ranged from 5 to 15 dB, indicating the utility of this source to transmit controllable infrasound signals over distances of 5 km.

PACS numbers: 43.28.Dm, 43.28.Hr, 43.38.Ja [RW]
Date Received: October 14, 2008 **Date Accepted:** January 8, 2009

1. Introduction

The ability to generate controllable infrasound has presented a technical challenge (Bedard and Georges, 2000); one reason is the mass-dominated radiation impedance found in conventional transduction technologies at very low frequencies. For example, the University of Mississippi designed and built a large horncoupled electropneumatic loudspeaker for the U.S. Army Research Laboratory known as the mobile acoustic source (Neill, 1993). The MOAS is frequently referred to as the "mother of all speakers" and was capable of projecting high intensity sound to frequencies as low as 10 Hz, though it was a large, complex system. Use of conventional loudspeaker technology in an array has demonstrated the ability to project acoustic signals down to frequencies of 8 Hz (Walker *et al.*, 2008), though likely not with sufficient sound pressure levels to be detected over ranges on the order of several kilometers.

Availability of a portable device that could project controllable infrasound detectable at ranges of several or tens of kilometers could find immediate use within the infrasound community. For example, the International Monitoring System (IMS) of the Comprehensive Test Ban Treaty Organization (CTBTO) (2001) operates a world-wide network of infrasound monitoring arrays. A portable, controllable source could enable field calibration and verification of operational arrays. Another important issue to the infrasound community is the evaluation of new sensors and the characterization of noise reduction systems, both of which could benefit from a controllable source. Other potential uses might include a surrogate for explosives that are currently employed to generate signals for testing of tactical explosion localization systems, or as a probe for atmospheric propagation studies.

The Thigpen rotary woofer (TRW) (Eminent Technologies, 2008) is a novel low frequency acoustic transducer capable of projecting controllable infrasound from a compact, energy efficient source (Garcés and Park, 2007). The TRW is essentially a baffled fan with blades that have dynamically controlled pitch. An electric motor rotates the hub at constant frequency while blade pitch is dynamically modulated by a control signal. The pitch control mechanism uses a conventional electromagnetic voice coil assembly driven by an audio amplifier to actuate the blades in response to the applied signal. The voice coil longitudinal oscillations are converted to blade axis rotary motion within the hub. Rotation of the blades about their axis changes the angle of attack of the blades as they move through the air creating uneven pressure distributions. It is believed that the TRW changes the radiation impedance from mass-

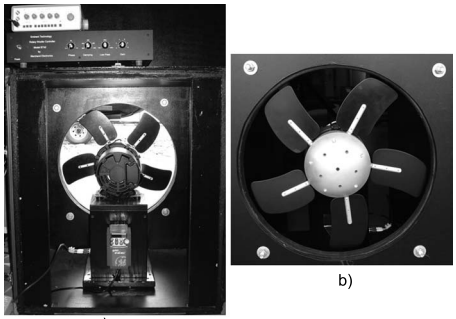


Fig. 1. Photograph of the TRW in a fixed installation. (a) View from inside the building showing the function generator, amplifier, motor controller, and motor, (b) view from the exterior showing the main rotor hub and blades.

dominated to fluid-displacement controlled by creating relatively high particle velocities at the blades of the projector thereby overcoming the canonical difficulty of mass-dominated radiation impedance at low frequencies.

The tested configuration consisted of five equally spaced blades, each approximately B_w=10.2 cm (4 in.) in width by B_L=15.3 cm (6 in.) in length. The outer diameter of the hub is approximately 20.4 cm (8 in.) so that the annular region of radiation has inner and outer radii of R_i=10.16 cm (4 in.) and R_o=25.40 cm (10 in.). Figure 1 shows a photograph of the TRW installed in a fixed laboratory.

Sections 2 and 3 outline portable deployment of the TRW, and demonstrate capability to transmit coherent infrasound signals detectable at ranges of 5 km.

2. Field deployment

To provide a field-deployable mobile prototype, the TRW was installed in a standard 17 ft U-Haul moving van. The installation was achieved by fashioning a baffle for the fan from 1.91 cm (3/4 in.) thick plywood. The fan housing is bolted to the baffle, and the baffle is clamped to the frame of the truck body. A portable generator and power-line filter were used to provide electric power.

On September 26, 2008 this portable system was deployed approximately 3.8 km from the central element ($H1$) of IMS array I59US on the western flank of Hualalai volcano, HI. A photograph of the portable system at the deployment site (19°34′38.32″N, 155°55′35.33″W) is shown in Fig. 2. An aerial view of the deployment layout is shown in Fig. 3. Table 1 lists the approximate distance from the source to each element of the I59US array. The array sensors are

Fig. 2. (Color online) The TRW in its portable configuration, located approximately 3.8 km from the central element of I59US.

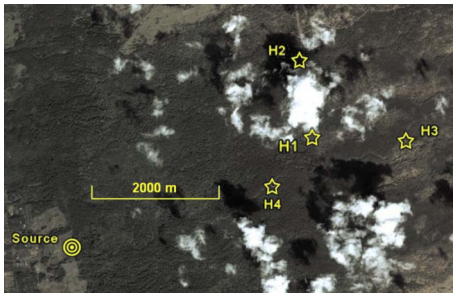


Fig. 3. (Color online) An aerial view of the deployment layout. The source was lcoated at approximately 19°34′38.32″N, 155°55′35.33″W.

Chaparral Physics (2008) Model 2.2A infrasound microphones. The data are digitized at a sample frequency of 20 Hz and recorded with GPS timestamps.

3. Results

A series of continuous wave (cw), constant frequency, and amplitude sine waves were broadcast according to the schedule listed in Table 2. The fan hub rotational speed for all transmissions was 12 Hz (720 rpm). Figure 4 plots the pressure power spectral density (PSD) recorded at the array sensors for broadcast numbers 2, 3, 8, and 9 of Table 2. These transmissions correspond to frequencies of 8, 6, 5, and 7 Hz, respectively. The three sensors, which are closer than 4.2 km (*H*1, *H*2, and *H*4), exhibit signal-to-noise ratios (SNRs) on the order of 15 dB at all transmission frequencies. Sensor *H*3 has noticeably reduced signal power at the range of 4.9 km. Even with the reduced SNR of approximately 5 dB at 4.9 km, the signal energy is clearly distinct from the noise.

Table 1. Approximate distance from the infrasound source to each of the array elements of I59US.

Array element	Distance (km)
*H*1	3.77
*H*2	4.16
*H*3	4.90
*H*4	3.02

Table 2. Schedule of signals broadcast during the test. All signals were constant amplitude and frequency sine waves. V_{in} is the rms voltage of the blade modulation signal.

No.	Broadcast time (UTC)	Frequency (Hz)	V_{in} (V_{rms})
1	9/29/08 23:53–23:56	8	6.8
2	9/29/08 23:56–23:58	8	10.5
3	9/29/08 23:59–24:03	6	13.1
4	9/30/08 00:03–00:07	4	11.9
5	9/30/08 00:10–00:15	3	18.0
6	9/30/08 00:15–00:20	1	25.6
7	9/30/08 00:20–00:24	3	13.2
8	9/30/08 00:24–00:29	5	16.4
9	9/30/08 00:29–00:34	7	16.0
10	9/30/08 00:34–00:39	8	17.6

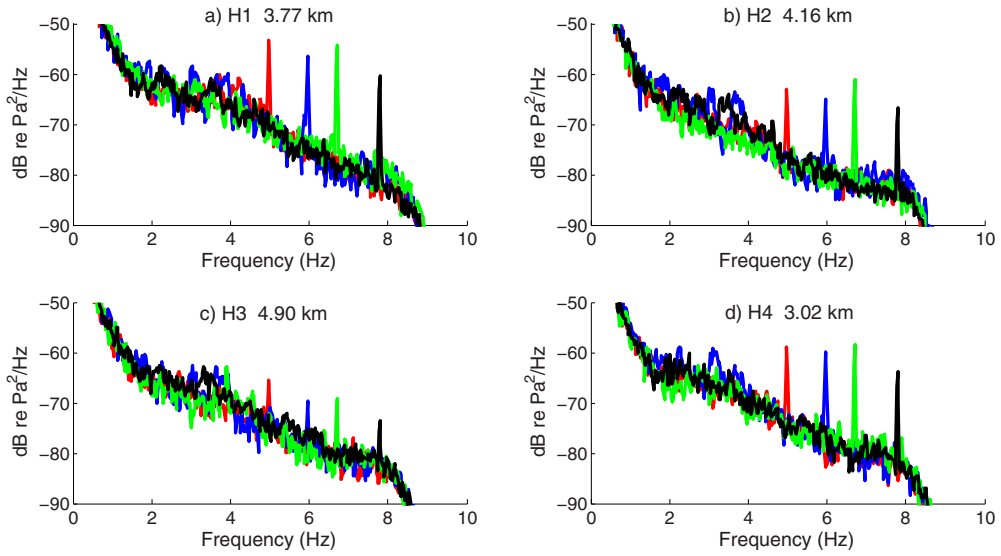


Fig. 4. (Color online) PSD of the recorded pressure at the four sensors from signal transmissions of 5, 6, 7, and 8 Hz. (a) Sensor *H*1, range 3.77 km; (b) sensor *H*2, range 4.16 km; (c) sensor *H*3, range 4.90 km; and (d) sensor *H*4, range 3.02 km.

The broadcast frequencies of 1, 3, and 4 Hz did not exhibit noticeable spectral peaks in the PSD. There are several possible reasons for this. One is the increase in noise levels below 5 Hz that could serve to mask the signals; another is that the output power of the source may attenuate at low frequencies due to the limited backvolume.

4. Conclusion

A prototype of the TRW modified for field deployment has demonstrated the ability to insonify an operational IMS infrasound array with cw signals from a stand-off distance of approximately 3.8 km. This may be the first controlled insonification of an operational IMS infrasound array from a portable device. SNRs in the range 5–15 dB indicate that this technology has merit as a field-deployable infrasound generator.

The development of a controllable infrasound generator will likely find utility in several acoustic applications. For example, the field calibration of infrasound arrays operated by the IMS of the CTBTO (2001), or as a tool for assessing atmospheric propagation studies. Other applications might include the field simulation of geophysical or anthropogenic infrasound sources to assist in the development of infrasound sensors and detection algorithms.

References and links

Bedard, B. J., and Georges, T. M. (**2000**). "Atmospheric infrasound," Phys. Today **53**(3), 32–37.
Chaparral Physics (**2008**). Geophysical Institute, University of Alaska, P.O. Box 757320, Fairbanks, AK 99775-7320.
CTBTO, (**2001**). *The Global Verification Regime and the International Monitoring System*, (CTBTO Preparatory Commission, Vienna International Centre, Vienna, Austria).
Eminent Technologies (**2008**). 225 East Palmer Street, Tallahassee, FL 32301.
Garcés, M., and Park, J. (**2007**). "A rotary subwoofer as an infrasonic source," 2007 Infrasound Technology Workshop, Tokyo, Japan, 13–16 November.
Neill, S. P. (**1993**). "Acoustical characteristics of a very low frequency sound source," MS thesis, Department of Physics and Astronomy, University of Mississippi, Oxford, MS.
Walker, K., Dzieciuch, M., Zumberge, M., and DeWolf, S. (**2008**). "A portable infrasonic sensor calibrator down to at least 8 Hz," 30th Monitoring Research Review on Ground-Based Nuclear Explosion Monitoring Technologies, Portsmouth, VA, 23–25 September.

Motion of the lips of brass players during extremely loud playing

Samuel Stevenson, Murray Campbell, and Seona Bromage
School of Physics, University of Edinburgh, Edinburgh, Lothian EH9 3JZ United Kingdom
samuel.stevenson@ed.ac.uk, d.m.campbell@ed.ac.uk, seona@ph.ed.ac.uk

John Chick
School of Engineering and Electronics, University of Edinburgh, Edinburgh, Lothian EH9 3JZ United Kingdom
john.chick@ed.ac.uk

Joël Gilbert
Laboratoire d'Acoustique de l'Université du Maine (UMR CNRS 6613), Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France
joel.gilbert@univ-lemans.fr

Abstract: When a brass instrument is played loudly, the energy level of the higher harmonics increases dramatically. The generally accepted explanation for this is non-linear steepening of the wavefront and generation of shock waves within the instrument bore. However, it has also been suggested that changes in the player's lip vibration could play a role in generating this "brassy" sound. To test this hypothesis, the dependence of lip opening-area on time has been measured for different dynamic levels in trombones and horns. Results suggest that the behavior of the open area does not change dramatically when the instrument enters the brassy regime.

PACS numbers: 43.75.Fg [TR]
Date Received: November 27, 2008 **Date Accepted:** January 19, 2009

1. Introduction

One striking feature of orchestral brass instruments is that the timbre is strongly dependent on dynamic level. At the loudest levels, the sound becomes distinctly "brassy" or "cuivré." This dramatic effect occurs due to a marked increase in the energy levels of the higher harmonics,[1,2] and brass musicians commonly use this timbral change as a form of musical expression. It is interesting to note that the effect is more dominant in some instruments than others; for instance, the trombone is generally considered to be "brassier" than the euphonium.

Musical tones are generated in brass instruments as a result of the self-sustained oscillation of the lips of the player.[3] Air flows from the lungs of the player into the mouthpiece of the instrument via the lips, which act as a valve modulating the airflow. This paper describes an experimental study of brass instrument playing at high dynamic levels, and examines the relationship between changes in the nature of the lip motion and the development of brassy timbre.

It is now generally acknowledged that the onset of the brassy regime is primarily a result of the non-linear propagation of the acoustic wavefront along the bore of the instrument, as clearly demonstrated by Hirschberg *et al.*[2] When high amplitude pressure oscillations are generated by the lips, the leading edge of the pressure wave steepens as it progresses through the bore of the instrument. If the air column of the instrument is sufficiently long—and the amplitude of the oscillations high enough—then the pressure rise becomes nearly instantaneous. This "shock-wave" formation gives rise to an increase in the higher frequency components in the tone which is produced.[4]

It has also been suggested by some researchers that the degree to which the lips can open becomes "saturated" or "clipped" during extremely loud playing as the movement of the lips becomes constrained by the rim and cup walls of the mouthpiece.[5–7] This effect, if present, would contribute significantly to the production of a brassy sound. Martin[8] performed some

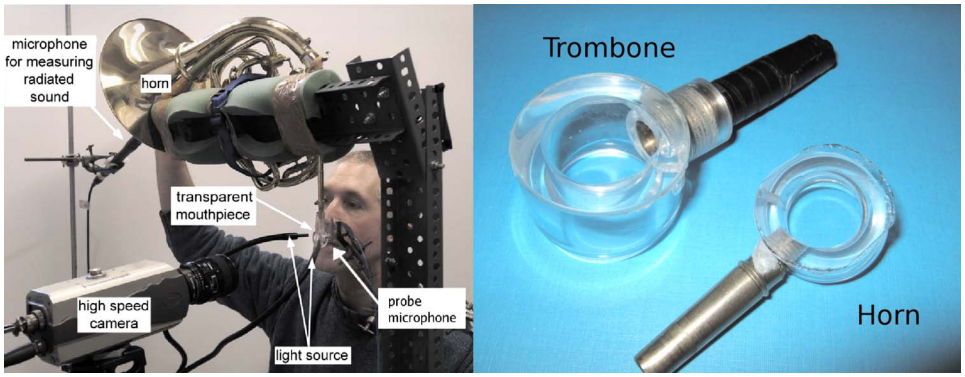


Fig. 1. (Color online) An overview of the complete experimental setup (left) for the horn showing the placement of the microphones, camera, and lightsource. The transparent mouthpieces (right) for both trombone and horn.

early measurements in which he showed that during mezzo forte playing on the cornet the variation of lip open area with time was almost sinusoidal. Other studies, using trombones, showed that the open area between the lips did not behave purely sinusoidally.[9,10] In fact, the "closing" phase of the lip motion took longer than the "opening" phase. This effect was more pronounced in large amplitude playing.

2. Experimental method

Three musicians—skilled amateurs with many years playing experience—were asked to sound pairs of notes at the same pitch but differing dynamic levels. One clearly brassy to the ear of the player (although not at the upper limit of loudness) and the other "non-brassy." The sound radiated from the bell of the instrument was captured using a Brüel and Kjær 4192 microphone at a distance of one bell radius. A PCB 106B pressure transducer was inserted flush with the inner surface of the trombone mouthpiece cup to capture the pressure waveform at the input to the instrument. For measurements of the input pressure in the (much smaller) horn mouthpiece, a short probe attachment was used with the PCB microphone.

The setup for the horn can be seen in Fig. 1 (left). During recordings on the horn, players placed their hand in the bell of the instrument as is normal for orchestral playing. Players were asked to keep their hand in the same position for both brassy and non-brassy playing. A previous study suggests that players are very consistent in where they place their hand.[11]

In order to allow optical access to the lips, transparent mouthpieces, as developed by Bromage,[12] were used [Fig. 1 (right)]. The key acoustic properties of the mouthpieces (throat, shank, cup volume, and rim dimensions) were based on current production designs in order to ensure that the mouthpieces are both realistic and comfortable for the player.

The motion of the lips was captured by a Phantom v.4 high speed digital camera recording at 6000 (trombone) or 8000 (horn) frames per second. At this speed, a strong light source is required and so a Schott model KL1500 cold light source was used to illuminate the lips. The captured footage was then split into a series of numbered grayscale bitmaps, one for each frame.

A threshold gray level was set. All pixels with a grayscale value above the threshold were set to white and all below the threshold (corresponding to the open area of the lips) set to black. This "binary" process can be seen in Fig. 2. For each image, the number of black pixels was counted and the lip open area as a function of time was deduced.

3. Results

Figure 3 shows examples of the radiated and mouthpiece pressure waveforms, recorded on a tenor trombone playing the notes F_3 and $B\flat_3$, for both brassy and non-brassy playing. Inspection of the mouthpiece pressure signals [Fig. 3 (left)] shows that the amplitude of the signal

Fig. 2. A sample lip image, including an example of the isolated open area.

increases as expected from non-brassy to brassy playing, but the waveforms are broadly similar. In contrast, comparison of the radiated sound pressures [Fig. 3 (right)] shows that there is a dramatic change in the form of the pressure signal between non-brassy and brassy playing. Non-linear propagation implies distortion of the pressure wave, which in turn can lead to the formation of a shock-wave. Here, there is clear evidence of shock-wave formation within the bore of the instrument, in agreement with the results of Hirschberg *et al.*[2] and also of Thompson and Strong.[13]

Analysis of the pressure signals recorded on the horn show similar behavior to that of the trombone, as can be seen from the online multimedia files Mm. 1 and Mm. 2. These are stereo recordings of the pressure signal in the mouthpiece (left channel) and the radiated sound (right channel) during brassy (Mm. 1) and non-brassy (Mm. 2) playing.

Mm. 1. A stereo recording of the pressure signal in the mouthpiece (left channel) and radiated sound (right channel) of a horn during brassy playing. Note C4. The starting transient has been removed. This is a file of type "wav." Size 258 kbytes.

Mm. 2. A stereo recording of the pressure signal in the mouthpiece (left channel) and radiated sound (right channel) of a horn during non-brassy playing. Note C4. The starting transient has been removed. This is a file of type wav. Size 258 kbytes.

The spectral centroid represents the distribution of power over frequency and provides a quantitative measurement of the increase in energy of the higher harmonics. It is defined as

$$F_{\mathrm{SC}} = \frac{\sum_{i=1}^{n} F_i A_i}{\sum_{i=1}^{n} A_i}, \tag{1}$$

where F_i is the frequency of the ith harmonic and A_i represents the amplitude of that harmonic.

The spectral centroids for both the mouthpiece pressure and the radiated sound during both brassy and non-brassy playing are shown in Table 1 for two notes recorded on the tenor trombone. For the note $\mathrm{B}\flat_3$ on the trombone, in the case of non-brassy playing, the spectral

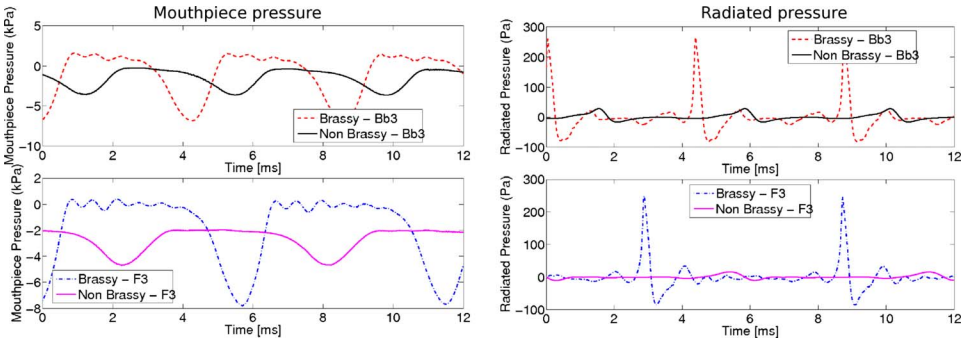


Fig. 3. (Color online) Mouthpiece (left) and radiated (right) waveforms for brassy and non-brassy playing. Notes $\mathrm{B}\flat_3$ (upper) and F_3 (lower) on the tenor trombone. Note that there is a dc offset of approximately 2 kPa in the mouthpiece pressure values.

Table 1. Spectral centroids (SCs) for both non-brassy (nb) and brassy (b) notes, calculated for both the mouthpiece pressure and the radiated sound.

		Mouthpiece pressure			Radiated pressure		
Instrument	Note	SC (nb) (Hz)	SC (b) (Hz)	SC(b)/SC(nb)	SC (nb) (Hz)	SC (b) (Hz)	SC(b)/SC(nb)
Trombone	$B\flat_3$	304	369	1.2	716	1816	2.5
	F_3	251	392	1.6	562	2462	4.4

centroid increases from 304 Hz in the mouthpiece to 716 Hz in the radiated sound—an increase by factor of 2.3. In the case of brassy playing, the centroid increases from 369 to 1816 Hz, an increase by factor of 4.9.

Part of the increase in spectral richness in loud playing is due to the high pass filtering effect of the bell. The effect of the linear filter of the bell at low amplitudes was estimated by comparing bell and mouthpiece frequency spectra. This filter was then applied to the measured mouthpiece spectrum for brassy playing to predict the bell spectrum if only linear propagation were significant. For the note F_3 on the trombone, we found that this calculated spectral centroid was 870 Hz, in comparison with the measured value of 2462 Hz. We conclude that the majority of spectral enrichment corresponds to non-linear propagation.

The spectral centroid of the sound recorded in the mouthpiece increased by a factor of 1.2 in the transition from non-brassy to brassy playing. In contrast, the ratio between the radiated spectral centroids is much larger—a factor of 2.5. The note F_3 as played on the trombone shows a similar trend. This suggests that the main source of the brassy sound does not come from within the mouthpiece of the instrument, but instead comes further downstream before the sound is radiated to the air. This quantitative analysis of the spectral centroid agrees with our qualitative visual analysis of the pressure waveforms shown in Fig. 3.

Comparison of mouthpiece and bell waveforms for the horn showed similar behavior as can be seen in online multimedia Mm. 1 and Mm. 2. However, a quantitative comparison of mouthpiece and bell spectra was not possible because of the difficulty of measuring the pressure in the small horn mouthpiece.

Figure 4 shows the lip open areas for both brassy and non-brassy playing for two notes on two different instruments—tenor trombone ($B\flat_3$ and F_3) (left) and horn (C_4 and F_3) (right). The brassy/non-brassy "pairs" are displayed with a common time-axis, but each lip open area

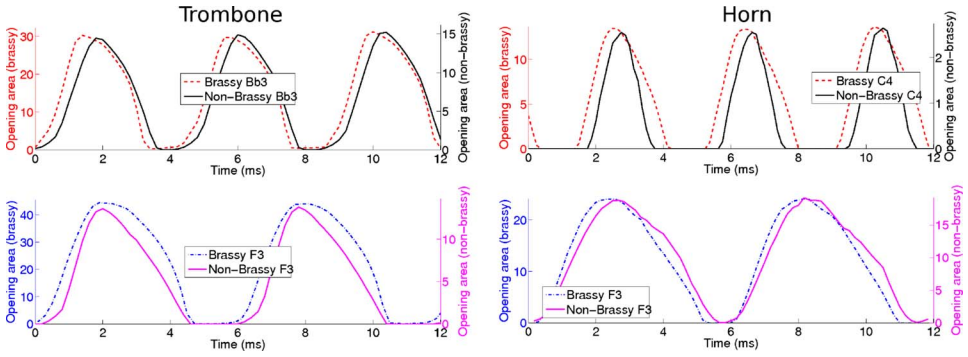


Fig. 4. (Color online) Lip open areas for brassy and non-brassy playing for two notes on the trombone (left) and two on the horn (right). All areas are given in mm^2. The data pairs share a common time-axis but each data set has its own *y*-axis, effectively normalizing the data for ease of comparison.

curve has its amplitude data displayed on a separate *y*-axis (brassy on the left, non-brassy on the right). Videos of both brassy and non-brassy playing corresponding to the note C4 on the horn are supplied as online multimedia in Mm. 3 and Mm. 4, respectively.

Mm. 3. A sample video of the lips of a horn player during brassy playing. Note C4. Framerate of 5 frames per second. This is a video of type "avi." Size 363 kbytes.

Mm. 4. A sample video of the lips of a horn player during non-brassy playing. Note C4. Framerate of 5 frames per second. This is a video of type avi. Size 363 kbytes.

A previous study, by Bromage,[12] found that the width of the lip opening increased quickly in the early part of the vibration cycle, and then remained at a constant value for much of the rest of the cycle. However, during this motion, the lip opening height varied continually and, accordingly, so did the lip open area. For louder notes, the width was constant for a greater proportion of the cycle than for quieter playing. We note, however, that it is the lip open *area* which is the primary variable by which the airflow into the instrument is controlled. From Fig. 4, it is immediately clear that the brassy/non-brassy pairs behave very similarly with respect to the lip open area. The only obvious difference between brassy and non-brassy playing is that, as expected, the brassy lip open areas have a larger amplitude than their non-brassy counterparts. For each brassy/non-brassy pair of measurements, the waveforms of the lip open area as a function of time are very similar, and do not show evidence of clipping in the brassy regime.

4. Conclusions

Spectral centroid data for the pressure signals in the mouthpiece and outside the bell of horns, and trombones support the hypothesis that the primary source of the increase in energy of the higher harmonics in very loud playing is to be found within the bore of the instrument. Studies of the lip motion of brass instrument players using transparent mouthpieces and a high speed digital camera show that the behavior of the lip open area does not vary significantly between brassy and non-brassy playing, providing confirmation that variations of the lip open area with amplitude are not the main source of the distinctive brassy timbre of brass instruments at the dynamic levels used by players in this study.

Acknowledgments

This work was funded by a Ph.D. grant (EP/P501407) from the EPSRC, U.K. The authors wish to thank all the musicians who offered their time and musical expertise.

References and links

[1]J. Backus and T. C. Hundley, "Harmonic generation in the trumpet," J. Acoust. Soc. Am. **49**, 509–519 (1971).
[2]A. Hirschberg, J. Gilbert, R. Msallam, and A. P. J. Wijnands, "Shock waves in trombones," J. Acoust. Soc. Am. **99**, 1754–1758 (1996).
[3]S. Elliott and J. M. Bowsher, "Regeneration in brass wind instruments," J. Sound Vib. **83**, 181–217 (1982).
[4]J. Gilbert, L. Menguy, and D. M. Campbell, "A simulation tool for brassiness studies," J. Acoust. Soc. Am. **123**, 1854–1857 (2008).
[5]G. Widholm, "The Vienna horn—A historic relic successfully used by top orchestras of the 21st century," in Proceedings of Forum Acusticum Budapest 2005, Budapest, Hungary (29 August–2 September 2005), pp. 441–445.
[6]T. R. Moore, E. T. Shires, I. E. W. Codery and A. Daniels, "The effect of bell vibrations on the sound of the modern trumpet," Acta. Acust. Acust. **91**, 578–589 (2005).
[7]N. H. Fletcher and A. Tarnopolsky, "Blowing pressure, power and spectrum in trumpet playing," J. Acoust. Soc. Am. **105**, 874–881 (1999).
[8]D. W. Martin, "Lip vibrations in a cornet mouthpiece," J. Acoust. Soc. Am. **13**, 305–309 (1942).
[9]D. C. Copley and W. J. Strong, "A stroboscopic study of lip vibrations in a trombone," J. Acoust. Soc. Am. **99**, 1219–1226 (1996).
[10]J. Chick, S. Bromage, D. M. Campbell, S. Stevenson, and J. Gilbert, "Motion of the brass player's lips during extreme loud playing," in Proceedings of the Eighth Congrès Français d'Acoustique, Tours (2006).
[11]J. Chick, C. Lumb, and M. Campbell, "Passive acoustic characteristics and intonation problems of modern orchestral horns," in Proceedings of the International Symposium on Musical Acoustics, Nara, Japan (2004).

[12]S. Bromage, "Visualisation of the lip motion of brass instrument players, and investigations of an artificial mouth as a tool for comparative studies of instruments,"Ph.D. thesis, University of Edinburgh, Edinburgh (2006).
[13]M. W. Thompson and W. J. Strong, "Inclusion of wave steepening in a frequency-domain model of trombone sound production," J. Acoust. Soc. Am. **110**, 556–562 (2001).

Estimating marine sediment attenuation at low frequency with a vertical line array

Yong-Min Jiang and N. Ross Chapman
School of Earth and Ocean Sciences, University of Victoria, P.O. Box 3055, Victoria, British Columbia V8W 3P6, Canada
mini@uvic.ca, chapman@uvic.ca

Kunde Yang and Yuanliang Ma
College of Oceanic Engineering, Northwestern Polytechnical University, Xi'an 710072, People's Republic of China
ykdzym@nwpu.edu.cn, ylma@nwpu.edu.cn

Abstract: This paper presents a method of determining the compressional wave attenuation in marine sediment from a short range measurement. The data were collected on a vertical line array at a range of 230 m during the Shallow Water 2006 experiments. The sediment attenuation is extracted from the signal strength ratio of the sea bottom reflection to a sub-bottom reflection at different frequencies from 1.75 to 3.15 kHz. Linear frequency dependence of the attenuation is found from the estimation. The sediment attenuation estimate is lower than the values estimated from the inversions of acoustic field data previously done in the vicinity.

PACS numbers: 43.30.Pc, 43.60.Pt [GD]
Date Received: December 9, 2008 **Date Accepted:** February 17, 2009

1. Introduction

Acoustic attenuation in marine sediment is an important parameter in understanding sound propagation in shallow water. Two approaches are generally used to estimate the sound attenuation in the sediment. One is based on *in-situ* measurements[1] and the other one is based on inferences from different kinds of inversions. *In-situ* measurements are usually done at very small ranges by measuring the signal levels between two embedded probes. The working frequencies are usually high, from several kilohertz to hundreds of kilohertz. Since the propagation distances are so small, this approach provides estimates of the intrinsic attenuation in marine sediments. In the second approach, the inversions that are used to infer sediment attenuation are usually applied to data collected at longer ranges. The frequencies are low, from tens to a few hundred hertz. Attenuation estimates from long range data inversions represent estimates of an overall loss parameter, which accounts for energy loss factors in addition to the intrinsic attenuation of the sediment material, such as sound scattering due to the roughness of the sea bottom and conversion to shear waves.

During the Shallow Water 2006 experiments (SW06),[2] short range experiments were carried out using a vertical line array (VLA) to estimate geoacoustic model parameters. This paper reports estimates of sediment attenuation from this data set. The method is based on measuring the differences of the signal levels from the bottom and sub-bottom reflections at different frequencies. The sediment layer thickness and *p*-wave sound speed are used to determine the values of sediment *p*-wave attenuation in dB/m and dB/λ.

2. Experiment and data

The experiment was carried out on 31 August 2006 on the New Jersey Shelf. The source was deployed from the research vessel Knorr to depths from 15 to 65 m in 10 m intervals. The data were collected on the Marine Physical Laboratory vertical line array (MPL-VLA1) located at (39°1.4771′N, 73°2.259′W). The range between the source and the receiver was approximately 230 m. The VLA consisted of 16 hydrophones equally spaced on the array with a sepa-

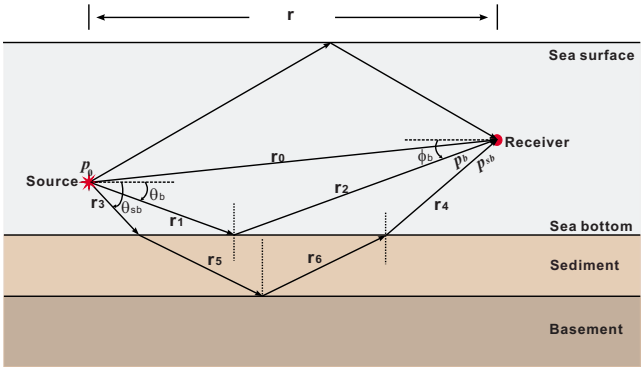


Fig. 1. (Color online) Geoacoustic model and the schematic of direct, surface reflected, bottom reflected, and sub-bottom reflected rays between the source and the receiver.

ration of 3.75 m, and the bottom most hydrophone (No. 1) was 8.2 m above the sea floor. The bathymetry between the source and MPL-VLA1 was almost flat with 79.6 m at the source and 79.0 at MPL-VLA1. The sound speed profiles (SSPs) in the water were recorded at the source position. A thermocline with a strong negative gradient was evident from 12 to 50 m of the SSP. The sediments in the vicinity of the site are clay-rich outer shelf materials.[3]

The signal transmitted was a 1-s linear frequency modulated pulse with a repetition period of 1 s. The signal frequency band was from 1500 to 4500 Hz. The direct, bottom reflected (BR), and surface reflected arrivals are clearly seen in the matched filtered signals for all of the source depths. Strong sub-bottom reflections from the ubiquitous "R-reflector"[3,4] are evident in the signals for most of the source-receiver pairs. The bottom reflected and sub-bottom reflected signals are relatively stable both in amplitude and arrival time. However, due to the oceanographic variation between the source and the VLA, strong fluctuations both in the amplitudes and in the arrival times of the direct and surface reflected arrivals are observed in the matched filtered signals. The complicated micro-structures in the thermocline are the main reason for multiple direct arrivals and consequently the fluctuation of the direct arrivals. As a result, the attempt of generating a measurement of bottom loss versus grazing angle described for an equivalent short range experimental geometry[5] was unsuccessful.

3. The sediment attenuation estimation method

Figure 1 is a schematic of the primary rays between the source and the receiver for the geometry in this study. Let p_0 be the pressure at the source, ξ the vertical directional response of the source, and ζ the vertical directional response of the hydrophone. ξ and ζ are usually functions of the frequency f, as well as θ, the sound launching angle from the source, and ϕ, the sound arrival angle at the receiver. The bottom reflected pressure at the receiver p_b, and the sub-bottom reflected pressure at the receiver p_{sb} are written as

$$p_b = \frac{p_0 V_{\mathrm{bott}} \xi_b(f,\theta_b)\zeta_b(f,\phi_b)}{r_1 + r_2} e^{-\alpha_w(r_1+r_2)}, \tag{1}$$

$$p_{\mathrm{sb}} = \frac{p_0 T_{ws} T_{sw} V_s \xi_{\mathrm{sb}}(f,\theta_{\mathrm{sb}})\zeta_{\mathrm{sb}}(f,\phi_{\mathrm{sb}})}{r_3 + r_4 + r_5 + r_6} e^{-\alpha_w(r_3+r_4)} e^{-\alpha_s^{(\mathrm{neper})}(f)(r_5+r_6)}, \tag{2}$$

where V_{bott} is the reflection coefficient at the sea bottom, V_s is the reflection coefficient at the sub-bottom interface, T_{ws} is the transmission coefficient of sound from the water into the sediment, T_{sw} is the transmission coefficient of sound from the sediment back to the water, r_1+r_2 is the sound path length for the bottom reflected arrival, r_3+r_4 is the sound path length for the

sub-bottom reflected arrival in the water, r_5+r_6 is the sound path length for the sub-bottom reflected arrival in the sediment, α_w is the attenuation in the water, and $\alpha_s^{(\text{neper})}(f)$ is the attenuation in the sediment in Np/m. Since the hydrophones are omni-directional, and the vertical dimension of the source is comparable to the wavelength within the frequency band we are interested in, the ratio of p_b to p_{sb} can be written as

$$\frac{p_b}{p_{\text{sb}}} = \left(\frac{V_{\text{bott}}}{T_{ws}T_{sw}V_s}\right)\left(\frac{r_3+r_4+r_5+r_6}{r_1+r_2}\right)e^{\alpha_w[(r_3+r_4)-(r_1+r_2)]}e^{\alpha_s^{(\text{neper})}(f)(r_5+r_6)}. \quad (3)$$

Equation (3) can further be expressed in logarithmic form as

$$20\log_{10}p_b - 20\log_{10}p_{\text{sb}} = 20\log_{10}B + \alpha_s^{(\text{neper})}(f)(r_5+r_6)(20\log_{10}e), \quad (4)$$

where $B=(V_{\text{bott}}/T_{ws}T_{sw}V_s)[(r_3+r_4+r_5+r_6)/(r_1+r_2)]e^{\alpha_w[(r_3+r_4)-(r_1+r_2)]}$. If we use $\alpha_s(f)$, which is in dB/m to replace $\alpha_s^{(\text{neper})}(f)(20\log_{10}e)$, then Eq. (4) becomes

$$20\log_{10}p_b - 20\log_{10}p_{\text{sb}} = 20\log_{10}B + \alpha_s(f)(r_5+r_6). \quad (5)$$

Several assumptions are made in this analysis. First of all, the value of α_w is very small, so its frequency dependence can be neglected in the analysis. Further, the frequency dependences of V_{bott}, V_s, T_{ws}, and T_{sw} are assumed to be negligible within the bandwidth and at the grazing angles at the bottom and the sub-bottom interfaces in this experiment. Consequently, the sediment p-wave attenuation in dB/m kHz, $\alpha_s^{(f)}$, can be obtained from the relationship of $20\log_{10}p_b - 20\log_{10}p_{\text{sb}}$ versus frequency. Further, the sediment attenuation in dB/λ is obtained using the sediment sound speed c_s in the formula $\alpha_s^{(\lambda)}=\alpha_s^{(f)}c_s/1000$.

4. Data processing procedure

For each source-receiver pair, the received signals were first matched filtered (i.e., compressed), and then the envelopes of the compressed pulses were extracted and aligned with respect to the sea bottom arrivals. Next, the aligned envelopes were averaged over 1 min to increase the signal to noise ratio. The means and the standard deviations of the amplitudes of the sea bottom reflection and sub-bottom reflection are used to control the data quality in order to carry out further analysis. Only the signals in which the sub-bottom and BR paths were clearly separated in time from other water column arrivals such as the surface reflection were used in the analysis.

To obtain the amplitude of the sea bottom and sub-bottom reflections at different frequencies, the compressed pulses were filtered by a series of band pass filters. The center frequencies and the bandwidths of those filters were designed to have (1) enough bandwidth to ensure adequate time resolution in the filtered band to distinguish the two arrivals, (2) reasonable frequency coverage to ensure the filtered signals in adjacent bands contain independent information, and (3) sufficient number of $(20\log_{10}p_b - 20\log_{10}p_{\text{sb}})$ versus frequency samples for the next step of the analysis. The center frequencies of the band pass filters used in this study were 1750–4150 Hz, in 200 Hz separations. The bandwidth of the filters was 500 Hz. Similarly, the envelopes of the band pass filtered signals were summed over 1 min to average out the signal fluctuations. Figure 2 is an example of the aligned compressed pulses and the averaged band pass filtered signal.

The amplitude differences of the sea bottom and sub-bottom reflections obtained at the frequencies up to 3150 Hz were included in the data analysis. For higher frequencies, the signal to noise ratio of the sub-bottom reflection was too low (lower than 2 dB). Both a linear and a power law frequency dependence of the sediment attenuation were investigated in the study. A least-squares method was used as the criterion of the goodness of fit of the observed data to each of the proposed dependences. The data indicated a linear frequency dependence of the attenuation. Figure 3 shows the values of $20\log_{10}p_b - 20\log_{10}p_{\text{sb}}$ obtained at each frequency band, for the source/receiver pair at 45/26 m. It is clear from Eq. (5) that the estimated slope of the line in Fig. 3 is the product of $\alpha_s^{(f)}$ and the sound travel path in the sediment r_5+r_6.

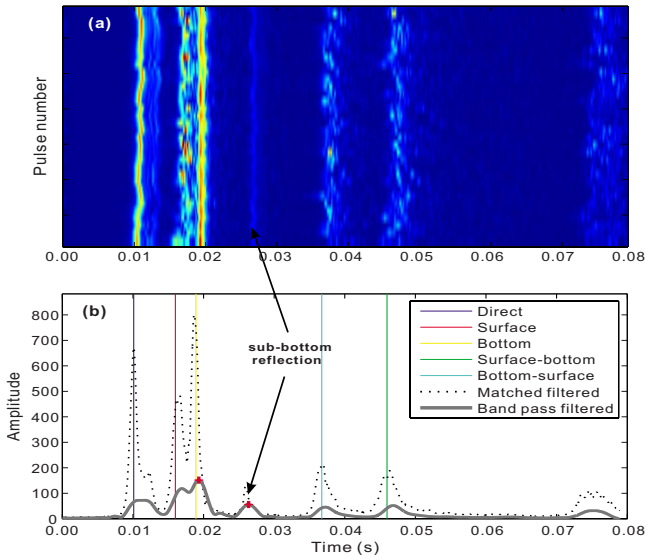


Fig. 2. (Color online) Pre-processed signals for source depth at 45 m and receiver depth at 26 m data. (a) Matched filtered signals for 1 min, plotted for a linear scale. Since only the relative values are used in the calculation, the absolute amplitude value is not specified; and (b) comparison of broadband matched filtered signal and filtered signal by a band pass filter centered at 1750 Hz.

The value of the sound path length in the sediments r_5+r_6 is derived from the sediment layer thickness and the incident angle at the sea bottom interface. The sediment sound speed and the sediment layer thickness were obtained from a multi-stage travel time inversion that has been reported previously.[6] First the water column SSP was inverted to account for the time varying water column environment, so that the effects of the water column variation are mitigated in the further analysis. The inverted profile was then used as an input in the inversion of the sediment sound speed and the layer thickness. Multiple hydrophone signals were used to constrain the inversion. The source launching angle, hydrophone receiving angle, and the incident angle at the sea bottom and sub-bottom interfaces were also obtained in the inversion.

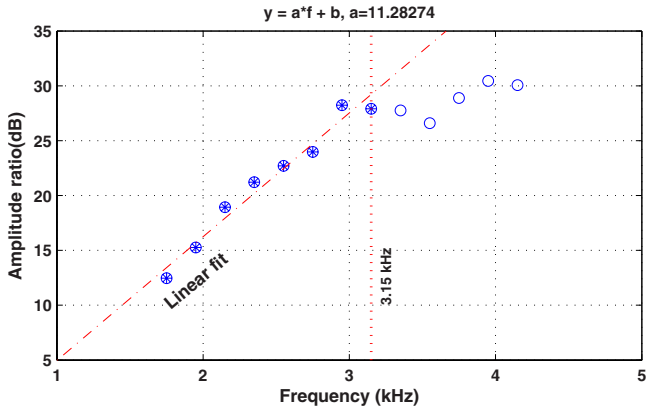


Fig. 3. (Color online) Data fitting of the sediment attenuation versus frequency. The circles are the differences of the amplitude of sea bottom and sub-bottom reflections obtained at different frequency bands. The stars are the actual data points used in the data fitting.

Table 1. Summary of marine sediment attenuation estimates.

Source depth (m)	35		45	
Receiver depth (m)	26.0	29.75	26.0	33.5
Sediment sound speed (m/s)[a]	1597		1600	
Sediment layer thickness (m)[a]	20.6		22.3	
Attenuation (dB/m kHz)	0.092	0.085	0.101	0.088
Attenuation (dB/λ)	0.15	0.14	0.16	0.14

[a]Reference 6.

5. Results and discussion

Table 1 lists the attenuation estimates from the data for four source-receiver pairs. For the four pairs of data, the sum of the standard deviations of the amplitudes of the sea bottom and the sub-bottom reflections is significantly smaller than the variations of $20 \log_{10} p_b - 20 \log_{10} p_{sb}$ from frequency to frequency. The grazing angles are greater than the critical angle at the sea floor and at the R-reflector.

The sediment attenuations are found to have linear frequency dependence in the frequency band of 1.75–3.15 kHz. The estimated sediment attenuations are from 0.085 to 0.101 in dB/m kHz, or from 0.14 to 0.16 in dB/λ. These estimates represent the average values in the layers. The uncertainties of the estimates are mainly from the signal fluctuations both in the amplitude and the travel time that are introduced by the variation of the source and receiver positions, the time varying ocean environment, and the inversion method.

The estimates reported here can be compared with other estimates from the vicinity. Choi *et al.*[7] analyzed close range high frequency (1–20 kHz) bottom loss data from the same site, and inferred the attenuation from a very thin surficial sediment layer from inversions of the measured loss. They found that a frequency dependence of $0.2(f/f_{ref})^{1.6}$ provided the best fit of the measured data at six grazing angles, where f_{ref} is 1 kHz. Turgut and Yamamoto[8] performed *in-situ* sound speed and attenuation measurements at the same site using a wideband acoustic probe, and found a linear frequency dependence of the attenuation for the band 10–80 kHz. Jiang and Chapman[9] inverted long range, multi-tonal low frequency (53–753 Hz) data from an experiment along the same radial path from MPL-VLA1, assuming a non-linear frequency dependence of the attenuation in the sediment layer down to the R-reflector. Their results indicated no sensitivity to attenuation in matched-field inversion for ranges of 1 and 3 km. However, at 5 km, sensitivity to attenuation increased due to the long range propagation distance, and they reported estimates of $0.6(f/f_{ref})^{1.4}$, where the reference frequency f_{ref} is 1 kHz. This result produces greater values for the range 2–3 kHz than the measurements reported here. Other estimates inferred from inversions have been reported by Carey and co-workers[10,11] from sites in the vicinity. Their values are also higher than the measurements reported here.

The estimates of attenuation from inversions of acoustic field data are affected by several different loss mechanisms, in addition to intrinsic attenuation in the sediment. The higher attenuations reported from the acoustic field inversions may simply reflect the effects of other attenuation mechanisms on the results of the experiments. Although the effects of other attenuation mechanisms cannot be ruled out in our experiment, the short range single bottom interaction geometry provides a closer approximation to the simple experimental geometries that are used successfully for measuring attenuation at high frequencies.

6. Conclusion

Marine sediment attenuation for a site on the New Jersey continental shelf was estimated from broadband acoustic data measured by a VLA. The short range–varying source depth–VLA geometry provided a practical way of estimating sediment attenuation at low frequencies. The estimated values are lower than the estimates of attenuation based on inversions of acoustic data.

Acknowledgments

This work is supported by the Office of Naval Research under Grant No. N00014-05-1-0264. The authors would like to thank Dr. William S. Hodgkiss and Dr. Peter Gerstoft for the acoustic data, and Dr. John Goff for the chirp seismic reflection data interpretation.

References and links

[1]M. D. Richardson and K. B. Briggs, "Relationships among sediment physical and acoustic properties in siliciclastic and calcareous sediment," ECUA2004, Delft (2004), Vol. **2**, pp. 659–664.
[2]D. J. Tang, J. Moum, J. Lynch, P. Abbot, R. Chapman, P. Dahl, T. Duda, G. Gawarkiewicz, S. Glenn, J. Goff, H. Graber, J. Kemp, A. Maffei, J. Nash, and A. Newhall, "Shallow Water '06—A joint acoustic propagation/ nonlinear internal wave physics experiment," Oceanography (Wash. D.C.) **20**, 156–167 (2007).
[3]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).
[4]A. Turgut, D. Lavoie, D. J. Walter, and W. B. Sawyer, "Measurements of bottom variability during SWAT New Jersey Shelf experiments," *Impact of Littoral Environmental Variability on Acoustic Predications and Sonar Performance* (Kluwer, Dordrecht, 2000), pp. 91–98.
[5]C. W. Holland, R. Hollet, and L. Troiano, "Measurement technique for bottom scattering in shallow water," J. Acoust. Soc. Am. **108**, 997–1011 (2000).
[6]Y.-M. Jiang, N. R. Chapman, and P. Gerstoft, "Short range travel time geoacoustic inversion with vertical line array," J. Acoust. Soc. Am. **124**, EL135–EL140 (2008).
[7]J. W. Choi, P. H. Dahl, and J. A. Goff, "Observation of the R reflector and sediment interface reflection at the Shallow Water '06 central site," J. Acoust. Soc. Am. **124**, EL128–EL134 (2008).
[8]A. Turgut and T. Yamamoto, "In situ measurement of velocity dispersion and attenuation in New Jersey Shelf sediments," J. Acoust. Soc. Am. **124**, EL122–EL127 (2008).
[9]Y.-M. Jiang and N. R. Chapman, "Bayesian geoacoustic inversion in a dynamic shallow water environment," J. Acoust. Soc. Am. **123**, EL155–EL161 (2008).
[10]S. M. Dediu, W. L. Siegmann, and W. M. Carey, "Statistical analysis of sound transmission results obtained on the New Jersey continental shelf," J. Acoust. Soc. Am. **122**, EL23–EL28 (2007).
[11]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).

Sand acoustics: The effective density fluid model, Pierce/Carey expressions, and inferences for porous media modeling

Kevin L. Williams
Applied Physics Laboratory, College of Ocean and Fishery Sciences, University of Washington, Seattle, Washington 98105
williams@apl.washington.edu

Abstract: Recently, Pierce and Carey [J. Acoust. Soc. Am. **124**, EL308–EL312 (2008)] presented a low frequency analysis of sound propagation in sand/silty sediments. Here, equivalent expressions are presented using a low frequency expansion of an unconsolidated version of Biot porous medium theory. The resulting expression for attenuation allows identification of the non-dimensional parameter β in the Pierce/Carey result in terms of physical parameters. The agreement of these two derivations motivates further analyses. The results imply that porous media propagation models that treat the medium's inertia via a single component approximation disregard a fundamental physical effect resulting from the relative inertia of the grains and fluid and are thus incomplete.

PACS numbers: 43.30.Ma, 43.20.Jr [JL]
Date Received: December 1, 2008 **Date Accepted:** February 18, 2009

1. Introduction

Studies of the constitutive equations for sound propagation in sand sediments are ongoing. Within this topic, the low frequency behavior of sound recently received focused attention by Pierce and Carey.[1] By examining locally averaged stresses and taking into account "the bowing of the fluid velocity profile in the interstices when the overall local region is being accelerated"[1] and the counteraction of this bowing via viscosity, they derived low frequency expressions for sound speed and attenuation. The attenuation they derived includes an explicit dependence on the difference between the density of solid grains and the fluid within the interstices of those grains. A key point is that this result is derived within the context of an unconsolidated sediment, i.e., within the Biot theory nomenclature, no frame rigidity.

In Sec. 2, results equivalent to those of Pierce and Carey are presented, derived from an unconsolidated version of Biot theory known as the effective density fluid model (EDFM).[2] Comparison of expressions allows identification of the non-dimensional parameter β in the Pierce/Carey result in terms of physical parameters from which a numerical range for β can be established. Within Sec. 2, high frequency expressions for attenuation and sound speed are also given.

In Sec. 3, the full EDFM as well as the high and low frequency expressions are examined as functions of frequency and relative grain/fluid density via plots of attenuation and sound speed. The physical interpretation of the plotted results is given and the inferences to porous media acoustics propagation modeling are discussed. Summary is presented in Sec. 4.

2. Pierce/Carey and EDFM expressions

The Pierce/Carey expression for low frequency sound attenuation in sand (with slightly different nomenclature) is

$$\alpha = \omega^2(\beta d^2/\eta)(1/2\rho c_{1f})(\chi_s^2\chi_f(\rho_s - \rho_f)^2). \quad (1)$$

Within Eq. (1), ω is the angular frequency, χ_f is the porosity (the fraction of the sand volume occupied by fluid), χ_s is the fraction of the sand volume occupied by solid (equal to $1-\chi_f$), ρ_f is the fluid density, ρ_s is the grain density, d is the mean grain diameter, and η is the fluid viscosity. Also within the expression, the following definitions apply:

$$\rho = \chi_s \rho_s + \chi_f \rho_f \tag{2}$$

and

$$c_{\mathrm{lf}} = \sqrt{\frac{B_{\mathrm{eff}}}{\rho}}, \tag{3}$$

where c_{lf} is the low frequency sound speed, ρ is the mass density, and B_{eff} is the effective bulk modulus of the sand given by

$$B_{\mathrm{eff}} = (\chi_s / B_s + \chi_f / B_f)^{-1}. \tag{4}$$

B_s is the bulk modulus of the grains and B_f is the bulk modulus of the fluid.

With the same definitions, the EDFM predicts that the (complex) sound speed is

$$c = \sqrt{\frac{B_{\mathrm{eff}}}{\rho_{\mathrm{eff}}}}, \tag{5}$$

where

$$\rho_{\mathrm{eff}} = \rho_f \frac{\alpha_t \chi_s \rho_s + \chi_f (\alpha_t - 1)\rho_f + i\chi_f \rho \eta F / \rho_f \omega \kappa}{\chi_f \chi_s \rho_s + (\alpha_t - 2\chi_f + \chi_f^2)\rho_f + i\chi_f \eta F / \omega \kappa}. \tag{6}$$

In addition to the previous definitions, α_t is the tortuosity, κ is the permeability, and F is a function given in Eq. (5) of the EDFM (Ref. 2) paper. The low frequency asymptotic value of F is unity and its high frequency expression is

$$F_{\mathrm{hf}} = (a_p \sqrt{\omega \rho_f / \eta})((1+i)/4\sqrt{2}), \tag{7}$$

where a_p is the "pore-size parameter,"[2] defined as $a_p = \sqrt{8\alpha_t \kappa / \chi_f}$.

A low frequency expansion of Eq. (5) (and using $k = \omega / c$ within the plane wave expression e^{ikr}) gives

$$\alpha_{lf} = \omega^2 (\kappa / \chi_f \eta)(1/2\rho c_{lf})(\chi_s^2 \chi_f (\rho_s - \rho_f)^2). \tag{8}$$

Comparison of Eqs. (1) and (8) shows that $\beta = \kappa / (\chi_f d^2)$. Note also that the low frequency limit of the EDFM real sound speed is that of Pierce/Carey, Eq. (3) (that equation is also known as Wood's equation). Typical parameter ranges for κ are 10^{-11}–10^{-10} m^2, sand grain diameters are roughly in the range of 0.1–1 mm, and porosities are in the neighborhood of 0.4. For those parameter ranges, β would have values from 2.5×10^{-5} to 2.5×10^{-2}. (The estimate for β in Pierce/Carey[1] was 1.3×10^{-3}.) Plots of Eq. (8) are included in Sec. 3.

It is also useful, for what follows, to give the high frequency sound speed and attenuation expressions derivable from the EDFM [and using Eq. (7)]. Sound speed becomes

$$c_{\mathrm{hf}} = \sqrt{\frac{B_{\mathrm{eff}}}{\rho_{\mathrm{hf}}}}, \tag{9}$$

where

$$\rho_{\mathrm{hf}} = \rho_f \frac{\alpha_t \chi_s \rho_s + \chi_f (\alpha_t - 1)\rho_f}{\chi_f \chi_s \rho_s + (\alpha_t - 2\chi_f + \chi_f^2)\rho_f}. \tag{10}$$

The attenuation is

Table 1. Parameter definitions and values used for the EDFM model.

Parameter	Value
Porosity (χ_f)	0.40
Mass density of sand grains (ρ_s)	2650 kg/m^3
Mass density of water (ρ_f)	1000 kg/m^3
Bulk modulus of sand grains (B_s)	3.6×10^{10} Pa
Bulk modulus of water (B_f)	2.25×10^{9} Pa
Viscosity (η)	0.001 kg/m s
Permeability (κ)	1.0×10^{-10} m^2
Tortuosity (α_t)	1.25

$$\alpha_{\mathrm{hf}}=\left(\sqrt{\frac{\omega\eta}{2\rho_f}}\right)\left(\frac{\chi_f a_p}{8c_{hf}\kappa}\right)\times\chi_s^2\chi_f(\rho_s-\rho_f)^2\times(\alpha_t\chi_s\rho_s+\chi_f(\alpha_t-1)\rho_f)^{-1}$$
$$\times(\chi_f\chi_s\rho_s+(\alpha_t-2\chi_f+\chi_f^2)\rho_f)^{-1}. \tag{11}$$

The implications of Eqs. (1)–(11) are examined as a function of frequency and grain density in Sec. 3 using the parameter values shown in Table 1 which are the same as given in Ref. 2.

3. Attenuation and sound speed results

Physical insight can be gained by examining the results above as the density of the grains goes to infinity. In particular, one sees the following:

$$\lim_{\rho_s\to\infty}\alpha_{\mathrm{lf}}\to\infty, \tag{12}$$

$$\lim_{\rho_s\to\infty}c_{\mathrm{lf}}\to 0, \tag{13}$$

$$\lim_{\rho_s\to\infty}\alpha_{\mathrm{hf}}=\left(\sqrt{\frac{\omega\eta}{2\rho_f}}\right)\left(\frac{\chi_f a_p}{8c_{\mathrm{hf}}\kappa\alpha_t}\right), \tag{14}$$

$$\lim_{\rho_s\to\infty}\rho_{\mathrm{hf}}=\rho_f\frac{\alpha_t}{\chi_f}, \tag{15}$$

$$\lim_{\rho_s\to\infty}c_{\mathrm{hf}}=\sqrt{\frac{\chi_f B_{\mathrm{eff}}}{\alpha_t\rho_f}}. \tag{16}$$

These asymptotic results indicate the following: (1) if, as grain density goes to infinity, there are frequencies low enough that the low frequency expansion applies, then no sound propagates within the medium since the attenuation is infinite and the sound speed zero; (2) if, as grain density goes to infinity, there are frequencies high enough that the high frequency expansion applies, then the attenuation and sound speed become insensitive to the grain density.

Before examining further the question of how the frequency range of validity of the low and high frequency approximations change as grain density increases, some further points can be made using the above results. Any model that uses the mass density in the equation for sound speed (e.g., Wood's equation) will predict that the sound speed goes to zero as grain density goes to infinity. Thus no sound would propagate across a layer of such a medium. This derives from ignoring the fundamental physical effect of inertia that causes the fluid to move relative to the grains (the first ramification of which shows up clearly in Eq. (1) via the dependence on the relative solid and fluid densities). This inertial effect in combination with the finite

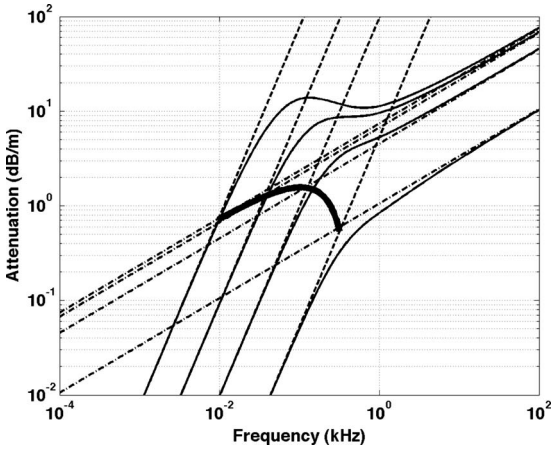


Fig. 1. The frequency dependence of the attenuation in a sand-like sediment. Results for the EDFM (solid lines), the low frequency expansion of the EDFM (dashed lines), and the high frequency expansion of the EDFM (dash-dot line). The parameter values used are given in Table I except that value of the grain density was varied by multiplication by 1, 4, 16, and 64 in going from the lowest to highest curves shown. The thick line segment indicates the value of attenuation at the frequency given by Eq. (17) as the grain density goes from that of sand to 64 times sand. That equation is an estimate of the high frequency limit of the low frequency expressions for attenuation and sound speed.

viscosity eventually results in the fluid breaking free of the solid as frequency is increased and thus causing attenuation and sound speed to become insensitive to grain density [i.e., Eqs. (14) and (16)]. That is, a model that includes this inertial effect (e.g., the Biot model) will predict that, regardless of the grain density of a porous medium, sound will propagate across a layer of that medium (at least at high frequency). Furthermore, it predicts that, as the grain density becomes large, the speed of this propagation is given by Eq. (16).

Now, the question remains as to the frequencies for which Eqs. (12)–(16) apply. Here, this question is examined graphically by using the full EDFM expression for sound speed and attenuation [derived from Eq. (5)] as well as the low and high frequency expressions [Eq. (8) or equivalently 1 and Eqs. (9) and (11)].

Figure 1 shows the attenuation of a porous medium as a function of frequency for the parameters given in Table 1 as well as for cases where the grain density is assumed to be 4, 16, and 64 times the density of sand. The figure shows the full model result and results for the low and high frequency expressions. This progression allows one to see how the frequency range of validity changes for the approximations. In particular, the high frequency approximation remains accurate above about 1 kHz for all grain densities and becomes insensitive to grain density for values above 16 times the density of sand. In contrast, the upper frequency limit of the low frequency approximation is reduced as the grain density increases. The implication is that the frequency range of accuracy for the low frequency approximation tends to zero as grain density goes to infinity. That is, though the asymptotic values given in Eqs. (12) and (13) are correct, the approximations leading to them are not.

Figure 2 shows the sound speed derived from the EDFM for the same parameter values used in Fig. 1. The low and high frequency approximations for each grain density are shown as squares near the ends of the frequency range. The high frequency, infinite density result [Eq. (16)] is shown as a line of diamonds. Again, one notes that as the grain density increases, the frequency range of validity for the low frequency approximation [Eq. (3), which is just Wood's equation] is reduced. This behavior implies that the frequency range of accuracy for the low frequency approximation for sound speed tends to zero as grain density goes to infinity.

From the expansions leading to the low frequency approximations for attenuation and sound speed, one can derive an estimate of the highest frequency (f_h) where the low frequency approximations may be accurate:

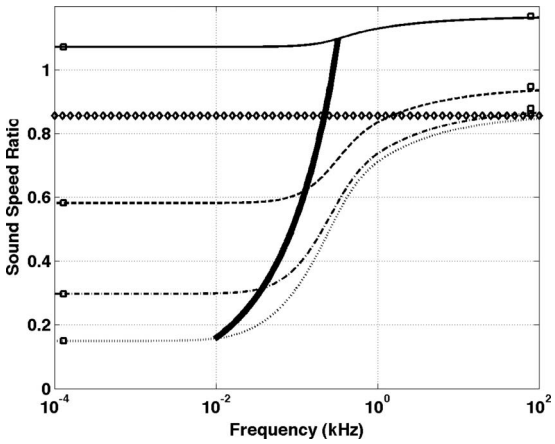


Fig. 2. The frequency dependence of the sound speed in a sand-like sediment (normalized by the sound speed in water as determined using the parameters for water given in Table I). Results for the EDFM are shown. The parameters values used are given in Table I except that value of the grain density was varied by multiplication by 1, 4, 16, and 64 for the solid, dashed, dash-dot, and dotted lines, respectively. The low and high frequency approximations for each grain density are shown as squares near the ends of the frequency range. The high frequency, infinite density result [Eq. (16)] is shown as a line of diamonds; comparison with the other curves indicates that at 100 kHz this limit has effectively been reached for grain densities of 16 times that of sand. The thick line segment indicates the value of sound speed ratio at the frequency given by Eq. (17) as the grain density goes from that of sand to 64 times sand. That equation is an estimate of the high frequency limit of the low frequency expressions for attenuation and sound speed.

$$f_h = \frac{0.1\chi_f \eta}{\kappa(\chi_f \chi_s \rho_s + (\alpha_t - 2\chi_f + \chi_f^2)\rho_f)}. \tag{17}$$

The values for attenuation and sound speed at f_h as the grain density is varied are shown as line segments in the figures. They indicate the f_h is a useful estimate. Examining f_h as the sediment density goes to infinity indeed shows that the useful range of the low frequency approximation goes to zero as grain density goes to infinity.

4. Summary

The low frequency expression for sound attenuation in a sand sediment derived by Pierce and Carey was compared to the low frequency limit of the Biot model in the limit that the medium is unconsolidated (i.e., the EDFM). This comparison allows one to identify the non-dimensional constant in the Pierce/Carey result in terms of physical parameters [i.e., $\beta = \kappa/(\chi_f d^2)$].

One of the important aspects of the Pierce/Carey derivation is that they show how to account for both the inertial effects due to the relative grain/fluid density and the effect of fluid viscosity in a relatively simple manner. The resulting expression for attenuation depends explicitly on the difference in grain and fluid densities. The agreement between the Pierce/Carey and EDFM low frequency attenuation expressions motivated the further analysis examining the high and low frequency limits of both attenuation and sound speed.

The general inference of the examination of attenuation and sound speed as a function of frequency and grain density is that any model that ignores this fundamental inertial effect is missing a key piece of physics that makes it incomplete. Wood's equation is one such model and since the Buckingham[3] model uses Wood's equation it is another example. The point being that, as shown in the analysis of Pierce and Carey and above, the use of the mass density to handle inertia in a porous medium is a low frequency approximation. Thus Wood's equation and Buckingham's model are low frequency approximations. However, even at low frequencies, incorrect treatment of inertia can manifest itself due to an incorrect dependence of attenuation on the relative densities of the grains and fluid. It will certainly manifest itself at high frequencies

because the density being used (mass density) is incorrect [i.e., Eq. (6) should be used]. One can formulate an incomplete model via introduction of hypothesized effects that influence the effective moduli. The moduli can thus be altered to accurately predict sound speed and attenuation as a function of frequency. However, this also masks the problem of incorrectly handling the inertial effect of relative grain/fluid motion. In essence, the problem lies in trying to correct an error in the resistance of the overall medium to motion [the inertia—the denominator in Eq. (5)] by adjusting its resistance to compression [the moduli—the numerator in Eq. (5)]. Though, perhaps mathematically possible, it is physically incorrect. One result of this error is that any prediction that requires acoustic impedance (e.g., the reflection coefficient[4–6]) will also be incorrect.

It is relatively easy to add the inertial effect to a model such as Buckingham's. An attempt to do so was carried out as part of the work by Hefner.[7] However, in doing so the conclusion was reached that one could match both attenuation and sound speed data without hypothesizing the strain-hardening effect put forth by Buckingham. Instead one could rely on a suggestion originally made by Biot to incorporate a complex, frequency dependent modulus. (It is notable that the lowest frequency sound speed data acquired during a recent ocean experiment were actually below Wood's equation prediction—Chotiros *et al.*[8] recently proposed an additional mechanism that could result in speeds below Wood's equation at low frequencies.)

It seems possible to experimentally test the predictions on the effect of relative grain/fluid density. A porous medium constructed using lead grains (whose density is over four times that of sand) represents one easily realizable upper limit while polystyrene grains (density slightly above water) represent another interesting limit.

From a general physics perspective, the results and discussion in Ref. 1 and the preceding sections show that Wood's equation is a low frequency approximation. The approximation essentially comes in treating the porous medium as a single component (from the inertial standpoint the fluid is always assumed to be "locked" to the grains, i.e., no relative motion) through the use of mass density. To correctly treat the inertia of the porous medium, the two components need to be addressed separately. Doing so leads to the prediction that attenuation depends on the relative densities of the two components. It also leads to different predictions for any quantity that requires the use of acoustic impedance since the effective density of this two component medium is not the mass density. One way to understand this one vs two component issue is to imagine a thought experiment where the grain density is taken to infinity. In the one component approximation, the sound speed goes to zero as grain density goes to infinity. On the other hand, if the grains and fluid are treated separately, the prediction is that there will always be a frequency range where sound will travel through the medium at a finite speed and with finite attenuation regardless of grain density. This result is counter to the prediction of any model that uses mass density to describe the inertia of the porous medium at all frequencies. More importantly, it is indicative of a fundamentally different treatment of the medium. In the end, inertial effects due to the two component nature of the medium need to be included before addressing additional physics that may also influence the frequency dependence of propagation within the medium.

Acknowledgments

The author thanks Dr. Darrell Jackson, Dr. Philip Marston, Dr. Eric Thorsos, Dr. Dajun Tang, Dr. Steve Kargl, and Dr. Brian T. Hefner for reading the first versions of this manuscript. Their comments and insights into the problem lead to a much improved manuscript.

References and links

[1]A. D. Pierce and W. M. Carey, "Low-frequency attenuation of acoustic waves in sandy/silty marine sediments," J. Acoust. Soc. Am. **124**, EL308–EL312 (2008).
[2]K. L. Williams, "An effective density fluid model for acoustic propagation in sediments derived from Biot theory," J. Acoust. Soc. Am. **110**, 2276–2281 (2001).
[3]M. J. Buckingham, "On pore-fluid viscosity and the wave properties of saturated granular materials including marine sediments," J. Acoust. Soc. Am. **122**, 1486–1501 (2007).
[4]H. J. Camin and M. J. Isakson, "A comparison of sediment reflection coefficient measurements to elastic and poro-elastic models," J. Acoust. Soc. Am. **120**, 2437–2449 (2006).

[5]N. P. Chotiros, in *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer, Dordrecht, The Netherlands, 1995), pp. 353–358.
[6]K. L. Williams, "Forward scattering from a rippled sand/water interface: Modeling, measurements and determination of the plane wave, flat surface reflection coefficient," IEEE J. Oceanic Eng. (in press).
[7]B. T. Hefner and K. L. Williams, "Sound speed and attenuation measurements in unconsolidated glass-bead sediments saturated with viscous pore fluids," J. Acoust. Soc. Am. **120**, 2538–2549 (2006).
[8]N. P. Chotiros, M. Isackson, and J.-X. Zhou, "Low-to mid-frequency model of attenuation and dispersion," J. Acoust. Soc. Am. **124**, 2468 (2008).

Statistical estimation of source location in presence of geoacoustic inversion uncertainty

Chen-Fen Huang
Institute of Oceanography, National Taiwan University, Taipei 10617, Taiwan
chenfen@ntu.edu.tw

Peter Gerstoft and William S. Hodgkiss
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238
gerstoft@ucsd.edu, wsh@mpl.ucsd.edu

Abstract: A statistical estimation of source location incorporating uncertainty in ocean environmental model parameters is derived using a Bayesian approach. From a previous geoacoustic inversion, a posterior probability distribution of the environmental parameters that reflects uncertainty in the ocean environment is obtained. This geoacoustic uncertainty then is mapped into uncertainty in the acoustic pressure field and is propagated through the Bartlett matched-field processor for source localization. Using data from the ASIAEX 2001 East China Sea experiment, the estimated source location and variability over time are compared with the known source positions.

PACS numbers: 43.30.Pc, 43.30.Wi, 43.60.Pt [JL]
Date Received: December 22, 2008 **Date Accepted:** February 18, 2009

1. Introduction

This paper presents a statistical estimation of source location in the presence of ocean acoustic parameter environmental uncertainty. Many Bayesian approaches for increasing robustness to ocean environmental uncertainty have been developed.[1–7] Averaging over a uniform prior distribution of environmental parameters was proposed earlier in Refs. 1 and 2 as a method of improving robustness of source localization. Dosso and Wilmut[5,6] proposed a Bayesian focalization to quantify source localization with environmental uncertainty. In their Bayesian focalization approach all source positions and environmental parameters are inverted at once. For practical problems involving thousands of source positions, tracking is an efficient alternative. Kalman and particle filtering[4,7] can be used efficiently to track the source and geoacoustic parameters.

Estimation of uncertainty in environmental parameters for a single source-receiver range is well developed using Markov chain Monte Carlo (MCMC) Metropolis–Hasting sampling.[8–10] The processor derived in this paper translates the uncertainty in environmental parameters, which, for example, can be from a previous geoacoustic inversion, to the uncertainty in replica fields for matched-field source localization. The resulting processor is a weighted sum of Bartlett processors over the posterior probability distribution (PPD) of the environmental parameters. We demonstrate the approach using source tow data from the ASIAEX 2001 experiment.[8,11] Using the geoacoustic inversion results for source localization is well known;[12] the focus here is on the accompanying source position uncertainty.

2. Theory

Figure 1 summarizes the statistical estimation of source position using environmental information obtained from a previous geoacoustic inversion.

The classical matched-field processor consists of systematically placing a single frequency test point source at each point of a search grid, computing the acoustic field (replica) at all the elements of the array and then correlating this modeled field with the observed data from

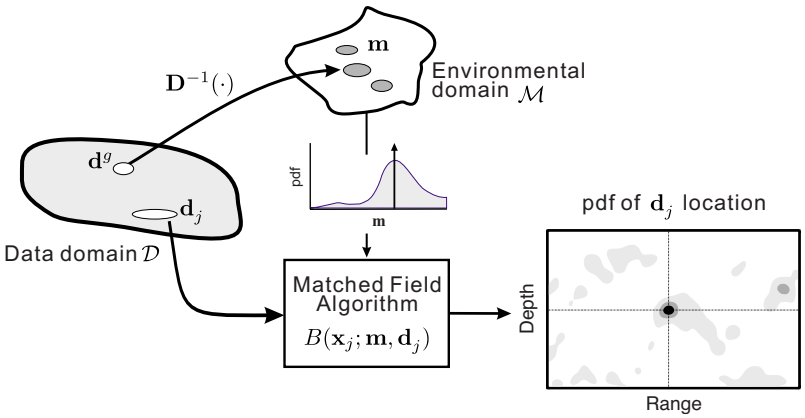


Fig. 1. (Color online) Measured data for source localization $\mathbf{d}_j$ ($\in\mathcal{D}$) is received at time j. From previous measured data $\mathbf{d}^g$ ($\in\mathcal{D}$) environmental parameters $\mathbf{m}$ ($\in\mathcal{M}$) are estimated via geoacoustic inversion along with a PPD of $\mathbf{m}$ that reflects its uncertainty. These environmental parameters then are mapped through a matched-field processor for statistically estimating the source location from $\mathbf{d}_j$.

an unknown source. In absence of data error and error in estimating the environmental parameter vector $\mathbf{m}$, the correlation is maximum when the test source location $\mathbf{x}$ is collocated with the true source

$$B(\mathbf{x}_j;\hat{\mathbf{m}},\mathbf{d}_j)=|\mathbf{w}(\mathbf{x}_j;\hat{\mathbf{m}})^H\mathbf{d}_j|^2=\mathbf{w}(\mathbf{x}_j;\hat{\mathbf{m}})^H\mathbf{R}_j\mathbf{w}(\mathbf{x}_j;\hat{\mathbf{m}}), \tag{1}$$

where $\mathbf{d}_j$ denotes the observed data at time index j, $\hat{\mathbf{m}}$ is an estimate of the parameter vector, and $\mathbf{w}(\mathbf{x};\hat{\mathbf{m}})=\mathbf{d}(\mathbf{x};\hat{\mathbf{m}})/|\mathbf{d}(\mathbf{x};\hat{\mathbf{m}})|$ is the normalized replica vector with the field solution $\mathbf{d}(\mathbf{x};\mathbf{m})$ computed from a acoustic model using the environmental parameter vector $\mathbf{m}$. When multiple snapshots are available, we assume that the signal source term can vary across snapshots, whereas the error power is the same for all snapshots. Then the observed data are efficiently expressed as a data covariance matrix $\mathbf{R}_j=\langle\mathbf{d}_j\mathbf{d}_j^H\rangle$.[13]

2.1 Posterior probability distribution of source position

The uncertainty in environmental parameters is translated to the uncertainty in replica fields for matched-field source localization.[8] The PPD of the source position at time j given the observation $\mathbf{d}_j$ is found via Bayes' rule

$$p(\mathbf{x}_j|\mathbf{d}_j)=\frac{p(\mathbf{d}_j|\mathbf{x}_j)p(\mathbf{x}_j)}{p(\mathbf{d}_j)}, \tag{2}$$

where $p(\mathbf{x}_j)$ is the probability density function (PDF) of $\mathbf{x}_j$ before observing the data. $p(\mathbf{d}_j|\mathbf{x}_j)$ is the likelihood function of the source position $\mathbf{x}_j$ at time step j given the data $\mathbf{d}_j$. This likelihood function can be derived from the PPD of $p(\mathbf{d}_j,\mathbf{m}|\mathbf{x}_j)$, by averaging over the environmental parameter vector $\mathbf{m}$ as follows:

$$p(\mathbf{d}_j|\mathbf{x}_j)=\int p(\mathbf{d}_j,\mathbf{m}|\mathbf{x}_j)\mathrm{d}\mathbf{m}=\int p(\mathbf{d}_j|\mathbf{m},\mathbf{x}_j)p(\mathbf{m}|\mathbf{x}_j)\mathrm{d}\mathbf{m}, \tag{3}$$

where $\mathbf{m}$ is independent of the source position $\mathbf{x}_j$ and the information about $\mathbf{m}$ was obtained from a previous geoacoustic inversion process using data $\mathbf{d}^g$. Therefore,

$$p(\mathbf{m}|\mathbf{x}_j)=p(\mathbf{m})=p(\mathbf{m}|\mathbf{d}^g). \tag{4}$$

The likelihood of the environmental parameters $\mathbf{m}$ and the source position $\mathbf{x}_j$ given $\mathbf{d}_j$ is determined from the assumption of data error statistics. The data error consists of both am-

bient noise as well as error in modeling the environment.[8] For simplicity, the error terms are assumed independent and identically distributed complex Gaussian random variables with variance ν as follows:

$$p(\mathbf{d}_j|\mathbf{m},\mathbf{x}_j,\nu)=\frac{1}{\pi^N\nu^N}\exp\left(-\frac{\phi(\mathbf{x}_j;\mathbf{m},\mathbf{d}_j)}{\nu}\right), \tag{5}$$

where $\phi(\mathbf{x}_j;\mathbf{m},\mathbf{d}_j)=\|\mathbf{d}_j\|^2-B(\mathbf{x}_j;\mathbf{m},\mathbf{d}_j)$ is the Bartlett objective function for the environmental parameter vector $\mathbf{m}$. Note that in Eq. (5) the complex source term has been estimated by a maximum likelihood approach[13] as opposed to integrating it out.[1] The uncertainty in the error variance ν is accounted for by integrating the complex Gaussian PDF over the nuisance parameter ν assuming a non-informative prior for ν (proportional to $1/\nu$) giving the likelihood function[11]

$$p(\mathbf{d}_j|\mathbf{m},\mathbf{x}_j)\propto\frac{1}{\phi(\mathbf{x}_j;\mathbf{m},\mathbf{d}_j)^N}, \tag{6}$$

where N is number of the hydrophones. Substituting Eqs. (6), (4), and (3) into Eq. (2) yields the PPD of $\mathbf{x}_j$ as

$$\begin{aligned}p(\mathbf{x}_j|\mathbf{d}_j)&=\frac{p(\mathbf{x}_j)}{p(\mathbf{d}_j)}\int p(\mathbf{d}_j|\mathbf{m},\mathbf{x}_j)p(\mathbf{m}|\mathbf{d}^g)\mathrm{d}\mathbf{m}\propto p(\mathbf{x}_j)\int\frac{1}{\phi(\mathbf{x}_j;\mathbf{m},\mathbf{d}_j)^N}p(\mathbf{m}|\mathbf{d}^g)\mathrm{d}\mathbf{m}\\&\propto p(\mathbf{x}_j)\sum_{k=1}^{K}\frac{1}{\phi(\mathbf{x}_j;\mathbf{m}_k,\mathbf{d}_j)^N},\end{aligned} \tag{7}$$

where $\phi(\mathbf{x}_j;\mathbf{m}_k,\mathbf{d}_j)$ is the Bartlett objective function for an environmental parameter vector $\mathbf{m}_k$ sampled by a MCMC algorithm. The statistical estimation of source position is based on Eq. (7).

3. Results

Data on a vertical line array from the ASIAEX 2001 East China Sea experiment in 106.5-m water depth are used to illustrate the approach.[8] First, an inversion was carried out at source-receiver range of 1.7 km, indicated by a circle in Fig. 2(a). An environment with 13 unknown parameters with their search bounds indicated in Fig. 2(b) was estimated, including geometrical, geoacoustic, and ocean sound speed empirical orthogonal function coefficients.[8] The inversion was based on a MCMC sampling giving samples $\mathbf{m}_k$ from which the PPD $p(\mathbf{m}|\mathbf{d}^g)$ could be estimated and subsequently used as prior information for source localization with new data $\mathbf{d}_j$. Based on Eq. (7), the samples $\mathbf{m}_k$ corresponding to a particular environment then are combined with the source position data $\mathbf{d}_j$ via the objective function $\phi(\mathbf{x}_j;\mathbf{m}_k,\mathbf{d}_j)$ for each possible location $\mathbf{x}_j$. The prior distribution $p(\mathbf{x}_j)$ was assumed uniform with the depth-range search interval. Summing over all samples gives the PPD of source location $p(\mathbf{x}_j|\mathbf{d}_j)$.

To estimate statistically the source position given the environmental information obtained from the inversion, a measured data vector $\mathbf{d}_j$ is used where the source range was 2.5 km. The source signal was a cw tone at 195 Hz. The signal-to-noise ratio (SNR) was estimated to be 16.3 dB using the ratio of the mean power received in the signal frequency fast Fourier transform bin and the (assumed to be noise) power received outside this bin. The observed data cross spectral density matrix $\mathbf{R}$ is normalized by the largest eigenvalue of the matrix, so that the maximum signal power is 1. For simulation purposes, the desired SNR is generated by varying the noise variance ν_n according to the array SNR definition[8]

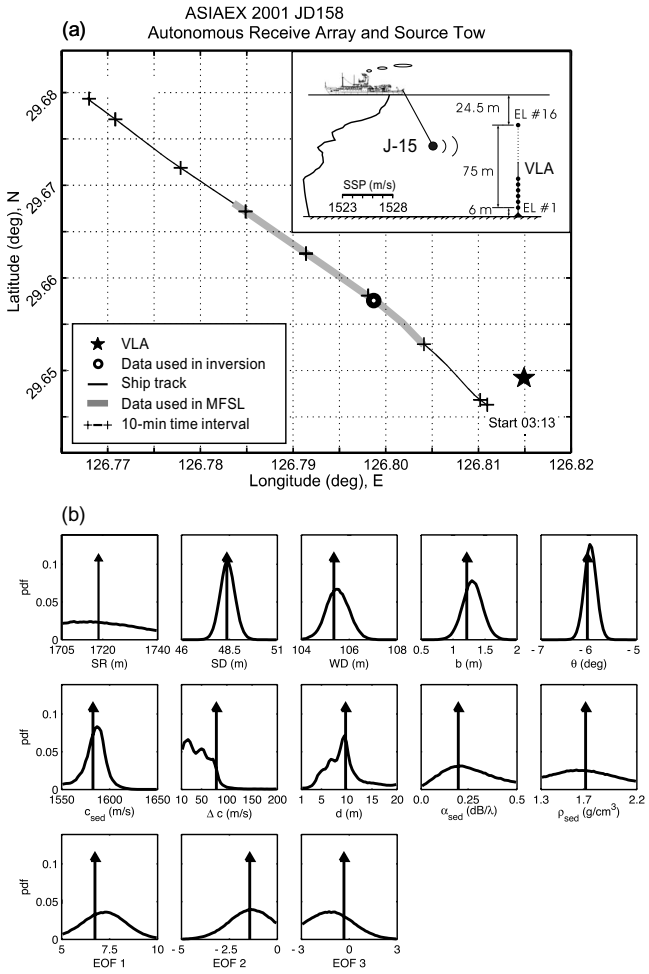


Fig. 2. (a) Track of R/V *Melville* during the ASIAEX 2001 East China Sea experiment. The experimental geometry is shown in the upper right-hand corner of the figure. (b) 1D marginal posterior probability densities and MAP (indicated by arrows) of the model parameters using the measured data obtained at approximately 1.7 km from the source.

$$\mathrm{SNR} = 10 \log \frac{1}{\nu_n} \quad \text{where } \nu_n = \frac{E[\|\mathbf{n}\|^2]}{N} \tag{8}$$

and $\mathbf{n}$ is the noise in the data (the array SNR should not be confused with the element SNR used in, e.g., Ref. 5). Since noise only can be added to the data, we only can lower the SNR from the maximum of 16.3 dB.

The parameter uncertainties are mapped through to the statistical estimation of source localization using Eq. (7). The replica vector for a hypothesized source position $\mathbf{x}$ and for each environmental parameter vector $\mathbf{m}_k$, $\mathbf{w}(\mathbf{x};\mathbf{m}_k)$, is computed using the normal mode propagation model SNAP.[14] Monte Carlo subsampling[8] of the geoacoustic inversion samples is used to evaluate Eq. (7) with $K=900$. The range-depth search grid spacing is 8 m in range from 1500 to 3500 m and 1.4 m in depth from 1 to 99 m.

Figure 3(a) shows the effect of the array SNR on the PPD of source range and depth at a source frequency of 195 Hz and range 2.5 km. The intersection of the lines indicates the peak of the PPD, and the darkest region indicates the 50% highest posterior density (HPD) in which

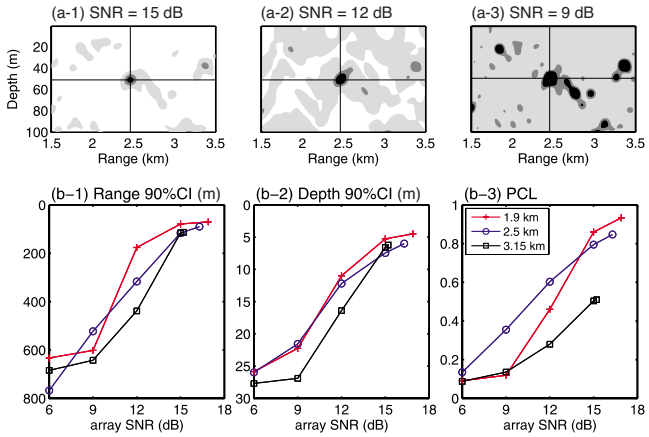


Fig. 3. (Color online) (a) Realizations of PPD of source range and depth for 195 Hz at 2.5 km range for three SNRs. Each panel shows the MAP source position (intersection of thin lines) and the HPD of source position [with 50% (dark), and 75% and 99% (light)]. (b) Source position uncertainty versus SNR for ranges 1.9 (+), 2.5 (○), and 3.15 km (□): (1) and (2) 90% CI, and (3) PCL in a region around the assumed source position.

the source falls given the observed data $\mathbf{d}_j$ and the environmental uncertainty. The range-depth PPD for the various SNRs [Figs. 3(a)(1)–3(a)(3)] shows that the number of competing peaks increases as SNR decreases. Note that the range-depth PPD for low SNR depends on the noise realization. The benefit here is not only we have the maximum *a posteriori* (MAP) estimate of source position but also the level of uncertainty.

The effect of SNR on the source position uncertainties is displayed for three source ranges in Fig. 3(b). The estimated SNRs in the observed data for ranges of 1.9, 2.5, and 3.15 km are 16.9, 16.3, and 15.2 dB, respectively (estimated using adjacent bins, see above). In Figs. 3(b)(1) and 3(b)(2) the 90% credibility interval (CI) extracted from the one-dimensional (1D) marginal distribution is shown. At lower SNR, the noise dominates the uncertainty of source position. For high SNR, the source position uncertainty will not approach zero since uncertainty in the environmental parameters will dominate. From the PPD surfaces in Fig. 3(a) we can

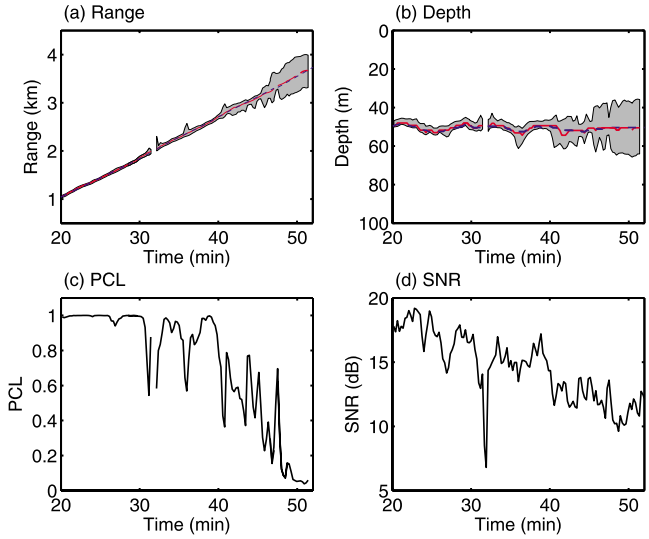


Fig. 4. (Color online) MAP estimate of source range (a) and depth (b) (solid) and their 90% CI (gray area) from time 20 to 52 min. The dashed line indicates the range measured by differential global positioning system (a) and the depth by a depth sensor (b). (c) PCL and (d) SNR estimated from the data.

integrate a region around the source position (an acceptable region of 20 m in depth and 200 m in range) obtaining the probability of correct localization (PCL) (Ref. 2) (this also has been termed integrated probability).[5] The PCL increases with SNR and tends to decrease with range.

As part of the experiment, a source tow from 1 to 3.7 km was performed. Using the estimated geoacoustic parameters and their uncertainty as well as the methodology above the MAP estimate of source position and the 90% CI is presented in Figs. 4(a) and 4(b). Furthermore, Figs. 4(c) and 4(d) show that as the SNR decreases so does the PCL.

4. Summary

Source position estimation in the presence of ocean environment uncertainty is investigated using the ASIAEX 2001 source tow data. A simulation shows that the noise deteriorates the localization performance at lower SNR, while the uncertainty in the ocean environment dominates at higher SNR. Continuous localization using the source tow data indicates that the localization uncertainty increases with range and 1/SNR.

Acknowledgments

This work was supported by the following: contract NSC96-2218-E-019-003 from the National Science Council of Taiwan; contract 96C100303 from the Ministry of Education of Taiwan under the project entitled Aim for Top University–Asian Pacific Ocean Research Center (to C.-F.H.); and Grant No. N00014-05-1-0264 from the Office of Naval Research.

References and links

[1]A. M. Richardson and L. W. Nolte, "*A posteriori* probability source localization in an uncertain sound speed, deep ocean environment," J. Acoust. Soc. Am. **89**, 2280–2284 (1991).
[2]J. A. Shorey, L. W. Nolte, and J. L. Krolik, "Computationally efficient Monte Carlo estimation algorithms for matched field processing in uncertain ocean environments," J. Comput. Acoust. **2**, 285–314 (1994).
[3]J. A. Shorey and L. W. Nolte, "Wideband optimal *a posteriori* probability source localization in an uncertain shallow ocean environment," J. Acoust. Soc. Am. **103**, 355–361 (1998).
[4]S. L. Tantum and L. W. Nolte, "Tracking and localizing a moving source in an uncertain shallow water environment," J. Acoust. Soc. Am. **103**, 362–373 (1998).
[5]S. E. Dosso and M. J. Wilmut, "Bayesian focalization: Quantifying source localization with environmental uncertainty," J. Acoust. Soc. Am. **121**, 2567–2574 (2007).
[6]S. E. Dosso and M. J. Wilmut, "Uncertainty estimation in simultaneous Bayesian tracking and environmental inversion," J. Acoust. Soc. Am. **124**, 82–97 (2008).
[7]C. Yardim, P. Gerstoft, and W. S. Hodgkiss, "Tracking of geoacoustic parameters using Kalman and particle filters," J. Acoust. Soc. Am. **125**, 746–760 (2009).
[8]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Validation of statistical estimation of transmission loss in the presence of geoacoustic inversion uncertainty," J. Acoust. Soc. Am. **120**, 1932–1941 (2006).
[9]Y.-M. Jiang, N. R. Chapman, and M. Badiey, "Quantifying the uncertainty of geoacoustic parameter estimates for the New Jersey shelf by inverting air gun data," J. Acoust. Soc. Am. **121**, 1879–1894 (2007).
[10]J. Dettmer, S. E. Dosso, and C. W. Holland, "Model selection and Bayesian inference for high-resolution seabed reflection inversion," J. Acoust. Soc. Am. **125**, 706–716 (2009).
[11]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Uncertainty analysis in matched-field geoacoustic inversions," J. Acoust. Soc. Am. **119**, 197–207 (2006).
[12]D. F. Gingras and P. Gerstoft, "Inversion for geometric and geoacoustic parameters in shallow water: Experimental results," J. Acoust. Soc. Am. **97**, 3589–3598 (1995).
[13]P. Gerstoft and C. F. Mecklenbräuker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," J. Acoust. Soc. Am. **104**, 808–819 (1998).
[14]F. B. Jensen and M. C. Ferla, *SNAP: The SACLANTCEN Normal-Mode Acoustic Propagation Model* (SACLANT Undersea Research Centre, SM 121, La Spezia, Italy, 1979).

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.

Monitoring stress related velocity variation in concrete with a 2×10^{-5} relative resolution using diffuse ultrasound (L)

Eric Larose[a)]
Laboratoire de Géophysique Interne et Tectonophysique, CNRS and Université J. Fourier, BP 53, 38041 Grenoble, France

Stephen Hall
Laboratoire 3S-R, CNRS and Grenoble Universities, 38041 Grenoble, France

(Received 19 June 2008; revised 12 January 2009; accepted 12 January 2009)

Ultrasonic waves propagating in solids have stress-dependent velocities. The relation between stress (or strain) and velocity forms the basis of non-linear acoustics. In homogeneous solids, conventional time-of-flight techniques have measured this dependence with spectacular precision. In heterogeneous media such as concrete, the direct (ballistic) wave around 500 kHz is strongly attenuated and conventional techniques are less efficient. In this manuscript, the effect of weak stress changes on the late arrivals constituting the acoustic diffuse coda is tracked. A resolution of 2×10^{-5} in relative velocity change is attained which corresponds to a sensitivity to stress change of better than 50 kPa. Therefore, the technique described here provides an original way to measure the non-linear parameter with stress variations on the order of tens of kPa.

[DOI: 10.1121/1.3079771]

PACS number(s): 43.20.Jr, 43.25.Ed, 43.25.Dc, 43.20.Gp [ROC] Pages: 1853–1856

Concrete is a complex heterogeneous mix of various ingredients of different sizes ranging from centimetric to millimetric gravel and sand to the micrometric porosity of the cement paste. Over the past 30 years or so, various sonic and ultrasonic techniques have been developed to image, characterize, or control the material.[1–4] Applications mostly work in the low frequency regime (a few Hz up to 50 kHz), where the wavelength is greater than the sizes of heterogeneities. In this regime, direct waves do not suffer too much from scattering by the heterogeneities and wave arrivals can be simply related to ballistic (direct or simply reflected) propagation paths. Nevertheless, because the wavelengths are much greater than the microstructure of the concrete, the fine details such as micro-damage or millimetric cracks are hardly visible. In the high frequency regime (above 50 kHz), the elastic wavefield is sensitive to these small details, but the ultrasonic wavelength is also of the order of the aggregates' size. This results in a strong attenuation of the direct (ballistic) wave,[5,6] and the appearance of a long lasting coda made of waves that have been multiply scattered by the internal microstructure.[7,8] The multiple scattering regime is often associated with a catastrophic loss of information, such that most conventional imaging or monitoring techniques fail. Nevertheless, as diffuse waves travel much longer paths than direct or simply reflected ones, they are much more sensitive to weak perturbation of the medium. This idea has been exploited in geophysics to quantify weak changes in the velocity of the earth crust:[9] by comparing two seismic coda obtained from fixed source and receiver at two different times, it is possible to monitor weak velocity variations in the medium. This technique is known as the *seismic doublet* technique or (more recently) *coda wave interferometry*.[10]

Various phenomena can cause acoustic velocity variations in concrete, including chemical reactions, and changes in water content, temperature, or stress. In this work, we focus on the latter effect, i.e., the relation between stress/strain and acoustic velocity, which is known as the acoustoelastic effect.[11–13] Under vertical uniaxial stress $\Delta\sigma$, the ultrasonic velocities in a sample will evolve as (to first order)

$$V_{ij} = V_{ij}^0 + \frac{\partial V_{ij}}{\partial \sigma}\Delta\sigma + o(\epsilon), \tag{1}$$

where ij stands for the wave velocity in direction i for a particle motion in j. The dimensionless non-linear parameter β is defined as

[a)]Electronic mail: eric.larose@ujf-grenoble.fr

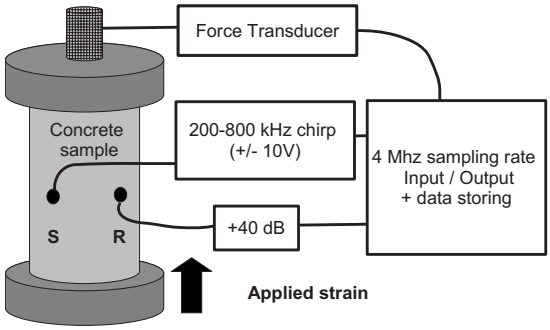


FIG. 1. Schematic view of the experimental setup.

$$\beta_{ij} = -\frac{E}{V_{ij}^0}\frac{\partial V_{ij}}{\partial \sigma} \tag{2}$$

and depends on the second (Lamé) and third (Murnaghan) order elastic constants. $E \approx 3.9 \times 10^{10}$ Pa is Young's modulus. There has been much less attention on the σ-dependence of V in heterogeneous media, such as concrete,[14] than in homogeneous media, and most results have considered variations under significant applied stress. To our knowledge, the use of diffuse ultrasound to evaluate the effect of stress loading on velocity variations has only been reported by Grêt *et al.*[15] and Payan.[16] In these works, the multiple time-window *doublet* technique was used to infer relative velocity variation with a relative resolution of $10^{-3}-10^{-4}$ for stress variation on the order of 1 MPa. Here, we propose to monitor weaker relative velocity variations in concrete down to 2×10^{-5} under weaker load variations (50 kPa). Such precision requires high frequency data (here 500 kHz diffuse ultrasonic data) and a more sensitive data processing technique than the *doublet* technique.

The experimental data presented here were acquired on a concrete sample of 16 cm in diameter and 28.5 cm in height. The concrete was prepared with 17% (in weight) of cement, 31% of fine sand, 43% of gravel (size ranging from 2 to 10 mm), and 9% of water. Two ultrasonic transducers are used in the experiment, one as a source and one as a receiver (labeled S and R, respectively, in Fig. 1). The lateral size of the active part of the transducers (0.7 mm) is much smaller than the wavelength, which makes them very sensitive to multiply scattered waves. The transducers are glued onto the sample using a hot chemical glue (phenyl-salicylic acid) that solidifies with cooling (below 43 °C). The ultrasonic experiment was performed several months after casting the sample.

During the ultrasonic measurements, the concrete specimen was subjected to uniaxial loading. A preliminary load was applied (5 MPa) at which point the reference ultrasonic data $r_0(t)$ were acquired. The applied uniaxial load was then increased to 5.5 MPa in ten steps of 50 kPa with ultrasonic acquisition at each stress increments, k, to provide $r_k(t)$. Each ultrasonic acquisition involved a source excited with a 10 V chirp $s(t)$ of frequencies ranging linearly from 200 to 800 kHz and of duration 10 s. The received signal $r_k(t)$ was amplified by 40 dB and stored in the computer. The recorded waveforms were subsequently cross-correlated with $s(t)$ (time-compression) to provide an estimate of the impulse response of the sample $h_k(t) = r_k(t)s(t)$, for each increment k.

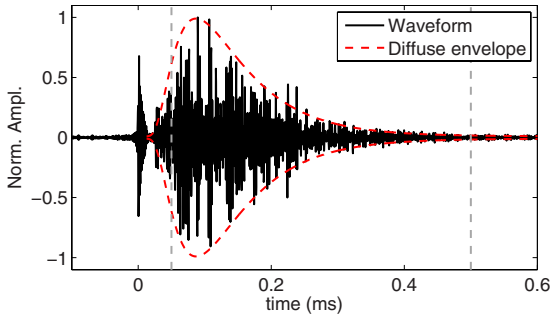


FIG. 2. (Color online) Typical waveform $h[t]$ collected through the concrete sample. The processed time-window marked by two vertical dashed lines extends from 50 to 500 μs.

The effective duration of the time-compressed signal $h_k(t)$ is about 0.6 ms. $h_0(t)$ is the reference waveform corresponding to $r_0(t)$ correlated with $s(t)$. It is noted that the source and receiver positions were kept fixed for all acquisitions and the excitation waveform $s(t)$ was not changed. Note also that the velocity change due to strain is one order of magnitude smaller than the one due to stress change, and is neglected here. The time separating two consecutive measurements is 5 min, over which no stress change was recorded and the ultrasonic records were reproducible. Creep was observed over much longer time scales (several hours) and was negligible in the present experiment.

A typical ultrasonic record is plotted in Fig. 2. This record is composed of (1) the impulse response of the concrete sample, (2) a peak at $t=0$ marking the initial excitation and due to the electronic cross-talk between S and R, and (3) background electronic noise (about 1% relative amplitude). The impulse response consists of a direct wave arriving at 0.023 ms, which is barely visible, and a following coda due to strong multiply scattered waves. The average intensity of the coda, $I(r,t)$, can be described by a diffusion envelope:

$$I(r,t) = \frac{I_0}{8(\pi D t)^{3/2}} e^{-r^2/4Dt - \gamma t}, \tag{3}$$

where I_0 is the energy released by the source, D is the diffusion constant, r is the source-receiver distance, and γ is the dissipation rate. In this experiment around 500 kHz, we find that $\sqrt{I}$ best fits the data for $D=5$ mm$^2/\mu$s and $\gamma=25$ ms^{-1}. This will represent a slight underestimate for D because Eq. (3) only holds for unbounded media. Nevertheless, D and σ values are in agreement with the literature,[7,8] and our rough estimation of D confirms that we are in the strong multiple scattering regime.

For each load k, the record $h_k(t)$ is compared to the reference waveform $h_0(t)$ to evaluate the relative velocity change in the concrete sample. Because of strong multiple scattering, the P- and S-wave arrivals mix together and rapidly attain the equipartition regime.[17] Let a_{ij} be the relative fraction of time that the wave spends in the state ij (polarization in the direction i while propagating in the direction j), then the velocity variation is $dV = \langle a_{ij}(\partial V_{ij}/\partial\sigma)\rangle \Delta\sigma$. Note that we have $\Sigma a_{ij}=1$ and that $\langle . \rangle$ denotes ensemble averaging. If the waves are statistically equipartitioned in all directions, we have $\Sigma a_{i=j}=9\%$ and $\Sigma a_{i\neq j}=91\%$.[17] In general, a_{ij} has a

more complex distribution, but in this work we focus on the average quantity $\langle a_{ij}\beta_{ij}\rangle$. For simplicity, we also assume that the waveforms are just stretched in time. This is again a first order approximation that applies to experimental data in the considered time-window. Nevertheless, velocity variations will progressively decorrelate the waveform.[18] This weak decorrelation, also called distortion, is not studied here although it contains precious information on the medium and its evolution. In particular, we stress that strong distortion and weak stretching might indicate a change in structure rather than a global velocity change. However, here we do not see significant distortion. To summarize, the velocity change dV for load k results in

$$h_k[t] = h_0[t(1+\nu_k)] + n(t), \tag{4}$$

with $\nu_k = dV/V$ and $n(t)$ accounting for additional fluctuations, including electronic noise and the small distortion mentioned earlier. Two processing techniques have been proposed in the literature to estimate dV/V. The first one, called the seismic *doublet* technique, was developed for geophysical purposes about 20 years ago.[9] The idea is to measure a time-shift between two different records in a limited time-window centered at time t. By measuring such time-shifts (or delay) δt at different times t in the coda, it is possible to evaluate the velocity variation, which is simply the average slope of $\delta t(t)$: $dV/V = -\delta t/t$. Such an approach implicitly assumes that the time-shift is constant within the considered time-window, which is generally not the case. Another idea[18,19] is to interpolate h_k at times $t(1-\nu)$ with various relative velocity changes ν.[20] Therefore, ν_k is the ν that maximizes the cross-correlation coefficient:

$$\mathrm{CC}_k(\nu) = \frac{\int_0^T h_k[t(1-\nu)]h_0[t]dt}{\sqrt{\int_0^T h_k^2[t(1-\nu)]dt \int_0^T h_0^2[t]dt}}. \tag{5}$$

Contrary to the *doublet* technique, we do not assume a constant time-shift in the considered time-window [0 T]. The comparison between the two waveforms can therefore be performed over the entire record at once, which results in a more stable, and thus more precise, estimation of dV/V. The comparison between the doublet and the stretching technique will be subject to further investigation.[21] Assuming that h_0 and h_i are stationary waveforms and are well described by Eq. (4), we have a theoretical estimation of CC:

$$\mathrm{CC}_k(\nu) = A\frac{\int_{\Delta f}\rho(f)\mathrm{sinc}(2\pi f(\nu-\nu_k)T)df}{\int_{\Delta f}\rho(f)df} + B(\nu), \tag{6}$$

with f the frequency, Δf the bandwidth, and $\rho(f)$ the spectral density. The constant A depends on the variance of h, noted $\langle h^2\rangle$ and the variance of the additional fluctuations, noted $\langle n^2\rangle$,

$$A = \frac{\langle h^2\rangle}{\langle h^2\rangle + \langle n^2\rangle}. \tag{7}$$

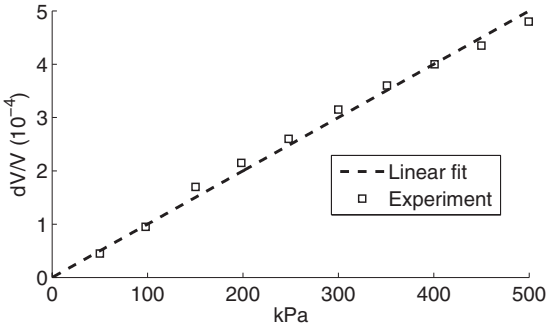


FIG. 3. Relative velocity variations dV/V versus stress variation $\Delta\sigma$.

$B(\nu)$ is a random process of zero mean and standard deviation:

$$\sqrt{\langle B^2\rangle} = \frac{1}{\sqrt{T\Delta f}}\frac{\langle n^2\rangle + 2\sqrt{\langle n^2\rangle\langle h^2\rangle}}{\langle h^2\rangle + \langle n^2\rangle}. \tag{8}$$

If the amplitude of the signal A is much greater than the fluctuations $\sqrt{\langle B^2\rangle}$ of the cross-correlation coefficient CC_k, the maximum of CC_k is obtained for the stretching $\nu = \nu_k$. It is interesting to note that the peak of the sinc function is visible even if the fluctuations (or noise) $n(t)$ are strong. In such a case, increasing the integration time T or the frequency bandwidth Δf can compensate for this strong noise. This is a crucial advantage of the present technique compared to the doublet technique. The dV/V resulting from our measurements are plotted in Fig. 3 versus $\Delta\sigma$. A linear regression of the data yields $\langle (1/V)(\partial V/\partial\sigma)\rangle = 10^{-6}$ kPa^{-1}, which gives an estimate of the non-linear parameter of our sample: $\beta = \langle a_{ij}\beta_{ij}\rangle \approx -40$. The standard deviation from the linear regression for the relative velocity change dV/V is 2×10^{-5}, which indicates that the precision on the evaluation of β is 5%. Note that during the whole experiment, the surrounding air temperature fluctuated within less than 0.8 °C. Temperature variations are too weak and thermal diffusion too slow in the sample (compared to the duration of the experiment) to induce noticeable dV.[10,18,22,23] This assertion is confirmed by the fact that the velocity variations dV are totally decorrelated from the air temperature variations. We therefore propose that the variations are due primarily to stress effects.

A potential application of this technique is on-site assessment of the non-linear constant β of a concrete structure. Our technique is sensitive enough to provide reliable measurements with additional stress $\Delta\sigma$ of the order of 100 kPa. As the non-linear parameter strongly depends on the micro-damage of the concrete, our technique would form an efficient and non-destructive way to probe the integrity of civil engineering structures. Another application is to evaluate the state of stress of a given structure with a predetermined β parameter and can be extended to geophysical media.[24–26]

ACKNOWLEDGMENTS

We are thankful to P. Roux, M. Campillo, S. Catheline, and R. Weaver for fruitful discussions and J.B. Toni for experimental help. This work was partially funded by UJF-TUNES and RNVOR programs.

[1]J. S. Popovics and J. L. Rose, "A survey of developments in ultrasonic NDE of concrete," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 140–143 (1994).
[2]O. Abraham, C. Leonard, P. Cote, and B. Piwakowski, "Time frequency analysis of impact-echo signals: Numerical modeling and experimental validation," ACI Mater. J. **97**, 647–655 (2000).
[3]D. M. McCann and M. C. Forde, "Review of NDT methods in the assessment of concrete and masonry structures," NDT & E Int. **34**, 71–84 (2001).
[4]C. Payan, V. Garnier, J. Moysan, and P. A. Johnson, "Applying nonlinear resonant ultrasound spectroscopy to improving thermal damage assessment in concrete," J. Acoust. Soc. Am. **121**, EL125–EL130 (2007).
[5]E. N. Landis and S. P. Shah, "Frequency-dependent stress wave attenuation in cement-based materials," J. Eng. Mech. **121**, 737–743 (1995).
[6]T. P. Philippidis and D. G. Aggelis, "Experimental study of wave dispersion and attenuation in concrete," Ultrasonics **43**, 584–595 (2005).
[7]P. Anugonda, J. S. Wiehn, and J. A. Turner, "Diffusion of ultrasound in concrete," Ultrasonics **39**, 429–435 (2001).
[8]J. Becker, L. J. Jacobs, and J. Qu, "Characterization of cement-based materials using diffuse ultrasound," J. Eng. Mech. **129**, 1478–1484 (2003).
[9]G. Poupinet, W. L. Ellsworth, and J. Frechet, "Monitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras fault, California," J. Geophys. Res. **89**, 5719–5731 (1984).
[10]R. Snieder, A. Grêt, H. Douma, and J. Scales, "Coda wave interferometry for estimating nonlinear behavior in seismic velocity," Science **295**, 2253–2255 (2002).
[11]F. D. Murnaghan, *Finite Deformation of an Elastic Solid* (Wiley, New York, 1951).
[12]D. M. Egle and D. E. Bray, "Measurement of acoustoelastic and 3rd-order elastic-constants for rail steel," J. Acoust. Soc. Am. **60**, 741–744 (1976).
[13]B. Mi, J. E. Michaels, and T. E. Michaels, "An ultrasonic method for dynamic monitoring of fatigue crack initiation and growth," J. Acoust. Soc. Am. **119**, 74–85 (2006).
[14]T. T. Wu and T. F. Lin, "The stress effect on the ultrasonic velocity variations of concrete under repeated loading," ACI Mater. J. **95**, 519–524 (1998).
[15]A. Gret, R. Snieder, and J. Scales, "Time-lapse monitoring of rock properties with coda wave interferometry," J. Geophys. Res. **111**, B03305 (2006).
[16]C. Payan, "Non destructive evaluation of concrete. Potential of non-linear acoustics," Ph.D. thesis, Université de la Méditéranée, Aix-En-Provence (2007).
[17]R. L., Weaver, "On diffuse waves in solid media," J. Acoust. Soc. Am. **71**, 1608–1609 (1982).
[18]O. I. Lobkis and R. L. Weaver, "Coda-wave interferometry in finite solids: Recovery of p-to-s conversion rates in an elastodynamic billiard," Phys. Rev. Lett. **90**, 254302 (2003).
[19]C. Sens-Schönfelder and U. C. Wegler, "Passive image interferometry and seasonal variations of seismic velocities at Merapi volcano, Indonesia," Geophys. Res. Lett. **33**, L21302 (2006).
[20]The interpolation of the data at times $t\,(1-\nu)$ uses a spline interpolation algorithm. It should be noted that this procedure performs well even with our 4 MHz sampling rate. In fact, through numerical tests, we have found that the numerical precision of our processing procedure depends mainly on the signal-to-noise of the data and only feebly on the sampling rate. However, even the effect of signal-to-noise is small; in the case of 1% of electronic noise the error of the estimated relative velocity change ν_k is found to be much smaller than 10^{-6}.
[21]C. Hadziioannou, E. Larose, P. Roux, and M. Campillo, "Stability of monitoring weak changes in multiply scattering media with ambient noise correlation: Laboratory experiments." In press.
[22]E. Larose, J. De Rosny, L. Margerin, D. Anache, P. Gouédard, M. Campillo, and B. Van Tiggelen, "Observation of multiple scattering of kHz vibrations in a concrete structure and application to monitoring weak changes," Phys. Rev. E **73**, 016609 (2006).
[23]V. Leroy and A. Derode, "Temperature-dependent diffusing acoustic wave spectroscopy with resonant scatterers," Phys. Rev. E **77**, 036602 (2008).
[24]K. Yamamura, O. Sano, H. Utada, Y. Takei, S. Nakao, and Y. Fukao, "Long-term observation of in situ seismic velocity and attenuation," J. Geophys. Res. **108**, 2317 (2003).
[25]F. Niu, P. G. Silver, T. M. Daley, X. Cheng, and E. L. Majer, "Preseismic velocity changes observed from active source monitoring at the Parkfield SAFOD drill site," Nature (London) **454**, 204–208 (2008).
[26]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).

Comment on "Measurement of the frequency dependence of the ultrasonic parametric threshold amplitude for a fluid-filled cavity" (L)

M. Strasberg[a)]
David Taylor Model Basin, NSWCCD, 9500 McArthur Boulevard, Bethesda, Maryland 20817-5700

(Received 23 January 2008; revised 31 October 2008; accepted 5 February 2009)

Measurements of the threshold amplitude of the pump signal for parametric excitation of ultrasonic standing waves in a water-filled container are described in the paper of Teklu *et al.* [(2006). J. Acoust. Soc. Am. **120**, 657–660]. The threshold is reported to *increase* with increasing pump frequency from 2 to 7 MHz, in apparent disagreement with earlier measurements made in 1970 over the same frequency range in similar apparatus showing a *decrease* in parametric threshold with increasing frequency. It is shown that the very theory presented in both papers may explain the frequency dependence observed with the new 2006 data rather than the old 1970 data if proper account is taken of the variation in the acoustic absorption coefficient of water with frequency. However, other questions raised by data in both papers indicate the possibility that the phenomenon they describe was *not* parametric excitation.
 [DOI: 10.1121/1.3089223]

PACS number(s): 43.25.Gf, 43.35.Yb [ROC] Pages: 1857–1859

I. INTRODUCTION

A paper published in 2006 (Ref. 1) describes new measurements, which apparently were intended to confirm and extend earlier measurements reported in 1970 (Ref. 2) made in similar apparatus and frequency range, both said to indicate parametric excitation of standing ultrasonic waves at megahertz frequencies in a small volume of liquid. Instead, however, the authors find the new measurements to be "in apparent discrepancy" with the measurements reported in the earlier paper.

Parametrically excited oscillations are oscillations that are not excited by a force or pressure applied directly to the excited modes but rather by a sinusoidal variation in the magnitude of a physical quantity (or "parameter"), which influences the natural frequency of the excited mode, the frequency of the parametrically excited oscillation differing from the frequency of the oscillating parameter. Parametric excitation requires the magnitude of the sinusoidal variation to exceed a value called the "threshold" value, and the oscillating parameter is conventionally called the "pump signal." The most easily excited mode, i.e., the mode with the lowest threshold, is usually a mode whose natural (or resonance) frequency is half the frequency of the pump signal.[3]

The main concern of the 2006 paper is the value of the threshold amplitude of the pump signal and its dependence on the pump frequency. For the measurements, standing ultrasonic waves are set up in a small container of water or other liquid between a rigid back plate and an X-cut piezoelectric quartz-crystal disk 1-in. in diameter. The pump signal is a sinusoidal voltage applied to the piezoelectric disk, which results in a small sinusoidal variation in its thickness and corresponding small oscillation in the separation distance between the disk and back plate. Since the natural frequency of a standing wave is inversely proportional to this distance, standing waves between the disk and back plate can be excited parametrically, in principle, if the voltage and corresponding oscillation in separation distance is large enough.

The main concern of the 2006 paper is the observed dependence of the threshold voltage for parametric excitation of standing waves on the frequency of the pump signal, which ranges from 2 to 7 MHz. The reported new 2006 data and the apparent discrepancy with the older 1970 data bring to mind several questions worthy of discussion, among them the following:

(1) The 2006 measurements indicate an increase in threshold amplitude with increasing pump frequency, whereas the 1970 measurements showed a *decrease* in threshold with increasing frequency although performed in similar apparatus and frequency range.
(2) Neither the 2006 nor the 1970 measurements show parametric excitation of a standing wave mode whose natural (or resonance) frequency is one-half the pump frequency, although this mode is usually the most easily excited.
(3) The measured threshold amplitudes are orders of magnitude smaller than the value predicted by the theory described in both papers.

II. THRESHOLD VARIATION WITH FREQUENCY

Figure 3 in the 2006 paper is a plot of the 1970 data (Fig. 5 in the 1970 paper) showing the measured threshold amplitude inversely proportional to frequency. Both papers contend that these data agree with a theoretical calculation of the threshold given by Eq. (12) in the 2006 paper, which, in the notation common to both papers but slightly modified in form, is

[a)]Electronic mail: murray.strasberg@navy.mil

$$A = l_0 \sqrt{\left[\frac{f_n - f_0}{f}\right]^2 + \left[\frac{a\alpha c}{2\pi f}\right]^2}, \quad (1)$$

where A is the threshold oscillatory variation in the distance between the face of the piezoelectric disk and back plate at the onset of parametric oscillations, l_0 is the mean distance between the disk and back plate, $2f$ is the pump *cyclic* frequency, f is half that frequency, f_n is the natural frequency of the parametrically excited oscillation, $a=(f_n/f)^2$, the Greek α is the sound absorption coefficient of the liquid at the excited frequency f_n in nepers per unit distance (1 Np equals 8.7 dB), and c is the ultrasound propagation velocity in the liquid.

For the old 1970 data, both papers assume that $f_0=f \approx f_n$, so the left term under the radical is zero or negligible compared with the right term and $a=(f_n/f)^2=1$. With these assumptions, the equation reduces to

$$A = \alpha c l_0 / 2\pi f, \quad (2)$$

which is Eq. (21) and Eq. (13), respectively, in the 1970 and 2006 papers. The authors of both papers interpret the presence of the frequency f in the denominator of Eq. (2) to indicate that the theoretical threshold is inversely proportional to frequency, therefore agreeing with the old 1970 data. But this conclusion is erroneous because it fails to account for the variation in the absorption coefficient with frequency. The absorption coefficients of water and most common liquids tend to increase with the square of frequency over the range of 1–10 MHz. Because of this variation, values of absorption coefficients at these frequencies are conventionally given in tables by the expression[4]

$$\alpha = K f_n{}^2, \quad (3)$$

where K is a constant whose value depends on the liquid (for water, $K=2.4\times10^{-16}$ s^2/cm) and f is cyclic frequency in hertz, with the subscript n added here to indicate that it is the frequency of the excited mode. Inserting α from Eq. (3) into Eq. (2), the latter becomes

$$A = c l_0 K f_n^2 / 2\pi f = c l_0 K f / 2\pi, \quad (4)$$

indicating that the theoretical threshold amplitude of the pump signal should *increase* proportionally with frequency—not decrease. Accordingly, this may explain the data shown in Fig. 4 of the 2006 paper rather than the 1970 data in Fig. 3.

An attempt to obtain a valid theoretical explanation of the 1970 measurements showing the threshold decreasing with increasing frequency requires re-examining the more general Eq. (1), which may now be written as

$$A = l_0 \sqrt{\left[\frac{f_n - f_0}{f}\right]^2 + \left[\frac{acKf}{2\pi}\right]^2}. \quad (5)$$

Only the left term under the radical can result in a threshold decreasing with frequency—and only if it is larger than the right term and its numerator does not increase with frequency faster than its denominator.

III. THE MISSING MODE

As stated in both papers, theory indicates that the mode with the lowest threshold is a mode whose natural frequency is one-half the pump frequency. However, the spectrum of the excited modes shown in Fig. 3 of the 1970 paper does not contain a line at half the 4 MHz pump frequency.

IV. ABSOLUTE VALUES OF THE THRESHOLD

Neither paper gives measured threshold values of the quantity A that can be compared directly with the theoretical values given by the above equations. Figure 4 of the 2006 paper plots threshold values for the new data in "arbitrary units," and Fig. 3 plots the old 1970 data as the "threshold drive voltage" applied to the piezoelectric disk. A threshold of about 300 V is shown for a pump frequency of 4 MHz. This voltage can be converted into an oscillating path length using the transduction conversion factor $(A/V)=3.2\times10^{-10}$ cm/V listed in Table 1 of the 1970 paper (a value in order-of-magnitude agreement with the X-cut quartz thickness-expander d constant of 2.3×10^{-10} cm/V).[5] Using the larger transducer conversion factor, the estimated measured threshold oscillation is $A=9.6\times10^{-8}$ cm.

Calculating the theoretical threshold based on Eq. (1) for comparison with the above measurement estimate results in a surprise. For the conditions applicable to the old 1970 data, the left term under the radical is neglected, and the right term is evaluated with $l_0=10$ cm, $a=1$, $c=1.45\times10^5$ cm/s, $f_n=f=2\times10^6$ Hz, and $\alpha=2.4\times10^{-16}f^2=9.6\times10^{-4}$ Np/cm $(8.4\times10^{-3}$ dB/cm) in accordance with Eq. (2). The surprising result for the theoretical threshold is $A=1.1\times10^{-4}$ cm—some three orders of magnitude larger than the estimated measured value. If the neglected left term in Eq. (1) were included, the theoretical value would be even larger.

The theoretical threshold should be four times larger for the new 2006 data at the same pump frequency because the factor a in Eq. (1) is stated to equal 4 for the data shown in Fig. 4. Although there is insufficient information to permit estimating the transducer oscillatory displacement for these data, it cannot be significantly larger than the estimated value for the 1970 data because of practical imitations in the voltage that can be applied to the quartz disk.

V. CONCLUSION

Assuming that the theoretical threshold value given by Eq. (1) and (4) and the estimated measured value are both valid, one is forced to speculate that the large disagreement between them may indicate that the observed phenomenon was not parametric excitation of standing waves—a speculation reinforced by the absence of the expected excited standing wave whose natural frequency is one-half the frequency of the pump signal.

The observed phenomenon may be associated with subharmonic generation due to nonlinear ultrasound propagation in the water as discussed, e.g., by Yen,[6] rather than with

parametric excitation. To indicate the possibility of nonlinearity, it would be useful to determine the sound pressure of the pump-frequency standing wave during future measurements.

[1]A. Teklu, M. S. McPherson, M. A. Breazeal, R. D. Hasse, and N. F. Declercq, "Measurement of the frequency dependence of the ultrasonic parametric threshold amplitude for a fluid-filled cavity," J. Acoust. Soc. Am. **120**, 657–660 (2006).

[2]L. Adler and M. A. Breazeale, "Generation of fractional harmonics in a resonant ultrasonic wave system," J. Acoust. Soc. Am. **48**, 1077–1083 (1970).

[3]See the second sentence of Background in Ref. 1 supra.

[4]K. F. Hertzfeld and T. A. Litovitz, *Absorption and Dispersion of Ultrasonic Waves* (Academic, New York, 1959), see Fig. 73-1 and Table 73-1.

[5]L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics* (Wiley, New York, 1962), see Table 12.1.

[6]N. Yen, "Experimental investigation of subharmonic generation in an acoustic interferometer," J. Acoust. Soc. Am. **57**, 1357–1362 (1975).

Time domain formulation of the equivalent fluid model for rigid porous media (L)

Olga Umnova[a] and Diego Turo[b]
Acoustics Research Centre, University of Salford, Salford M5 4WT, England

(Received 14 October 2008; revised 15 January 2009; accepted 26 January 2009)

A set of equations has been derived which corresponds to the time domain formulation of the equivalent fluid model. It models the propagation of an acoustic pulse in rigid frame porous material accounting for both viscous and thermal effects. It has been shown analytically and confirmed numerically that the equations can be reduced to those published previously in the limit of long and short duration pulses. Numerical solutions have been found for different pulse durations and the results have been compared with other time domain models.
 [DOI: 10.1121/1.3082123]

PACS number(s): 43.55.Ev, 43.20.Bi, 43.20.Jr [RMW] Pages: 1860–1863

I. INTRODUCTION

The equivalent fluid model originally formulated in 1987[1] is based on a set of independently measurable material parameters and satisfies physically correct behavior in both high and low frequency ranges. The model was later extended[2,3] to include thermal losses in a porous material. In its most general formulation it relies on the knowledge of six parameters, i.e., material porosity ϕ, its tortuosity α_∞, steady state flow resistivity σ_0, thermal permeability k', and finally viscous and thermal characteristic lengths Λ and Λ'. Even if it does not provide the highest level of accuracy for all materials and in all frequency ranges,[4] it still remains widely accepted as one of the most successful semiempirical models in acoustics of porous materials.

Both low and high frequency approximations of the equivalent fluid model can be analytically formulated in the time domain[5,6] and can be used to model propagation of long and short acoustic pulses, respectively. Recently it has been shown that the relaxational model[7] also allows analytical time domain representation.[8–10] Although quite accurate for many materials, this model requires different values of viscous and thermal relaxation times to be used in low and high frequency ranges to satisfy physically correct limits of complex density and complex compressibility functions.[1–3]

In this work, we provide a time domain formulation of the full equivalent fluid model without any restrictions on the frequency bands.

II. TIME DOMAIN MOMENTUM CONSERVATION AND CONTINUITY EQUATIONS

We start with the following frequency domain equations for pressure p and particle velocity v of the plane sound wave:

$$-i\omega\rho_0\alpha(\omega)v = -\partial_x p, \tag{1}$$

$$-i\omega C(\omega)p = -\rho_0 c^2 \partial_x v, \tag{2}$$

here ω is angular frequency, ρ_0 is the equilibrium density of air and c is the sound speed. In the framework of the equivalent fluid model the complex density $\alpha(\omega)$ and complex compressibility $C(\omega)$ can be formulated as

$$\alpha(\omega) = \alpha_\infty\left(1 + \frac{\sqrt{1 - i\omega\tau_2}}{-i\omega\tau_1}\right), \tag{3}$$

$$C(\omega) = \gamma - \frac{\gamma - 1}{1 + \frac{\sqrt{1 - i\omega T_2}}{-i\omega T_1}}, \tag{4}$$

where

$$\tau_1 = \frac{\alpha_\infty\rho_0}{\phi\sigma_0}, \quad \tau_2 = \tau_1\frac{M}{2}, \tag{5}$$

and

$$T_1 = \frac{N_{\mathrm{Pr}}\rho_0 k'}{\phi\eta}, \quad T_2 = T_1\frac{M'}{2}, \tag{6}$$

which can be interpreted as two pairs of viscous and thermal relaxation times, γ is adiabatic constant, η is the dynamic viscosity of air, N_{Pr} is Prandtl number, and

$$M = \frac{8\eta\alpha_\infty}{\phi\sigma_0\Lambda^2} \quad M' = \frac{8k'}{\phi\Lambda'^2} \tag{7}$$

are viscous and thermal shape factors.

The complex tortuosity function (3) can be rearranged in the following way:

$$\alpha(\omega) = \alpha_\infty\left(1 + \frac{1}{-i\omega\tau_1\sqrt{1 - i\omega\tau_2}} + \frac{\tau_2}{\tau_1}\frac{1}{\sqrt{1 - i\omega\tau_2}}\right).$$

Now applying the inverse Fourier transformation to Eq. (1), the following time domain momentum conservation equation can be easily derived:

$$\rho_0\alpha_\infty\partial_t V + I_0(t) = -\partial_x P, \tag{8}$$

where

[a]Author to whom correspondence should be addressed. Electronic mail: o.umnova@salford.ac.uk
[b]Electronic mail: d.turo@pgr.salford.ac.uk

$$I_0(t) = \frac{\rho_0 \alpha_\infty}{\tau_1}\sqrt{\frac{\tau_2}{\pi}}\int_{-\infty}^{t} \frac{\exp\left(-\frac{t-t'}{\tau_2}\right)}{\sqrt{t-t'}} \times\left(\partial_{t'}V(t') + \frac{V(t')}{\tau_2}\right)dt',$$

V and P are particle velocity and acoustic pressure in the time domain. In the derivation, Eq. (8) from Ref. 8 has been used. Expression (4) for the complex compressibility can be rearranged as

$$C(\omega) = 1 + (\gamma - 1)\frac{\sqrt{1 - i\omega T_2}}{\frac{T_1}{T_2}(\sqrt{1 - i\omega T_2})^2 + \sqrt{1 - i\omega T_2} - \frac{T_1}{T_2}}.$$

In the fraction above, the denominator can be factorized as $T_1/T_2(\sqrt{1-i\omega T_2}-x_1)(\sqrt{1-i\omega T_2}+x_2)$, where $x_{1,2} = \sqrt{(T_2/2T_1)^2+1} \mp (T_2/2T_1)$. Further algebra leads to the final expression for $C(\omega)$,

$$C(\omega) = 1 + \frac{(\gamma-1)T_2}{T_1(x_1+x_2)}\left(\frac{x_1}{\sqrt{1 - i\omega T_2} - x_1} + \frac{x_2}{\sqrt{1 - i\omega T_2} + x_2}\right).$$

Now using slightly modified version of Eq. A.2 from Ref. 10, inverse Fourier transformation of Eq. (2) can be performed which results in the following:

$$\partial_t P + I_1(t) + I_2(t) + I_3(t) = -\rho_0 c^2 \partial_x V, \tag{9}$$

where

$$I_1(t) = \frac{\gamma-1}{T_1}\int_{-\infty}^{t} \partial_{t'}P(t')\frac{\exp\left(-\frac{t-t'}{T_2}\right)}{\sqrt{t-t'}}dt',$$

$$I_2(t) = \frac{2(\gamma-1)x_1^2}{T_1(x_1+x_2)}\int_{-\infty}^{t} \partial_{t'}P(t')\exp\left((x_1^2-1)\frac{t-t'}{T_2}\right)dt'.$$

$$I_3(t) = -\frac{\gamma-1}{T_1(x_1+x_2)}\int_{-\infty}^{t} \partial_{t'}P(t')\sum_{m=1}^{2} x_m^2 \times\exp\left((x_m^2-1)\frac{t-t'}{T_2}\right)\operatorname{erfc}\left(x_m\sqrt{\frac{t-t'}{T_2}}\right)dt'.$$

Equations (8) and (9) constitute time domain formulation of the equivalent fluid model.

Now we shall demonstrate that Eqs. (8) and (9) can be reduced to those published previously in the limit of long and short duration pulses. The low frequency approximation of the equivalent fluid model[5] should be applicable to pulses with duration much longer than both τ_2 and T_2. This implies that their central frequencies are much lower than $1/2\pi\tau_2$ and $1/2\pi T_2$. The low frequency approximations of complex density and complex compressibility functions[5] are

TABLE I. Pulse parameters.

Regime	ω_0 (rad/s)	t_0 (s)	f_0 (Hz)
Long pulse (spheres)	50	0.1	11.25
Medium pulse (gravel)	100	0.04	22.50
Medium pulse (spheres)	7 000	0.000 65	1575.60
Short pulse (spheres)	20 000	0.000 3	4501.60

$$\alpha(\omega) \approx \alpha_\infty\left(1 + \frac{1}{-i\omega\tau_1}\right), \quad C(\omega) \approx \gamma. \tag{10}$$

The first equation can be obtained assuming $\omega\tau_2 \to 0$, the second requires $\omega T_2 \to 0$.

In the limit considered, $\partial_t V \ll V/\tau_2$ and V varies much slower with time than $\exp(-(t-t')/\tau_2)$. Consequently, the convolution integral $I_0(t)$ in Eq. (8) can be approximated as

$$I_0(t) \approx \frac{\rho_0\alpha_\infty}{\tau_1}\frac{V}{\sqrt{\pi\tau_2}}\int_{-\infty}^{t}\frac{\exp\left(\frac{t-t'}{\tau_2}\right)}{\sqrt{t-t'}}dt' = \frac{\rho_0\alpha_\infty}{\tau_2}V.$$

This results in the following low frequency/long pulse limit of the momentum conservation equation:

$$\rho_0\alpha_\infty\partial_t V + \frac{\rho_0\alpha_\infty}{\tau_1}V = -\partial_x P, \tag{11}$$

which agrees with Eq. (15) from Ref. 5.

The integrals $I_{1,2}(t)$ in the continuity equation (9) can be represented as

$$I_{1,2}(t) \approx a_{1,2}\partial_t P, \tag{12}$$

where $a_1 = (\gamma-1)T_1/T_1$ and $a_2 = 2(\gamma-1)x_1/(x_1+x_2)$. In the derivation, one of the identities $x_{1,2}^2 - 1 = \mp(T_2/T_1)x_{1,2}$ was used.

To estimate $I_3(t)$ in Eq. (9), the integral identity $\int_0^\infty e^{\beta x}\operatorname{erfc}(\sqrt{\alpha x})dx = 1/\beta(\sqrt{[\alpha/(\alpha-\beta)]}-1)$, which is equivalent to Eq. (6.283) from Ref. 11, has been used, assuming $\alpha = x_m^2/T_2$, $\beta = (x_m^2-1)/T_2$, and $x = t-t'$ which leads to the following:

$$I_3(t) \approx \frac{(\gamma-1)}{(x_1+x_2)}(x_1(x_1-1) + x_2(x_2-1))\partial_t P. \tag{13}$$

Summing Eqs. (12) and (13), the low frequency/long pulse approximation of Eq. (11) is derived:

$$\gamma\partial_t P = -\rho_0 c^2 \partial_x V. \tag{14}$$

This equation is equivalent to the second of Eq. (15) in Ref. 5.

The high frequency approximation of the dynamic density and compressibility functions can be obtained from Eqs. (3) and (4) assuming $\omega\tau_2 \gg 1$ and $\omega T_1 \gg 1$, $\omega T_2 \gg 1$:

$$\alpha(\omega) \approx \alpha_\infty\left(1 + \frac{\sqrt{\tau_2}}{\sqrt{-i\omega\tau_1}}\right), \quad C(\omega) \approx 1 + \frac{(\gamma-1)\sqrt{T_2}}{\sqrt{-i\omega T_1}}.$$

In this regime pulse duration is assumed to be much shorter than τ_2, T_1, and T_2. The convolution integral $I_0(t)$ in Eq. (8) can be simplified assuming that $V/\tau_2 \ll \partial_t V$ and that

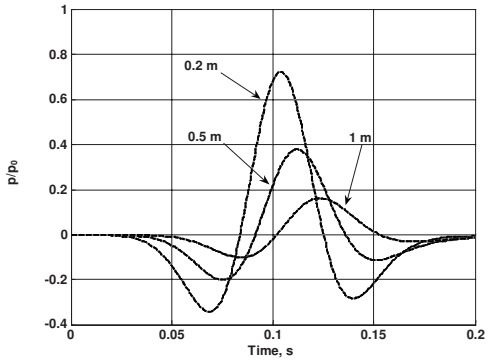


FIG. 1. Long pulse propagation in the packing of spheres. Solid line—equivalent fluid model (8) and (9), dotted line—low frequency approximation (11) and (14), and dashed line—relaxation model (Ref. 8) with relaxation times defined by Eq. (17).

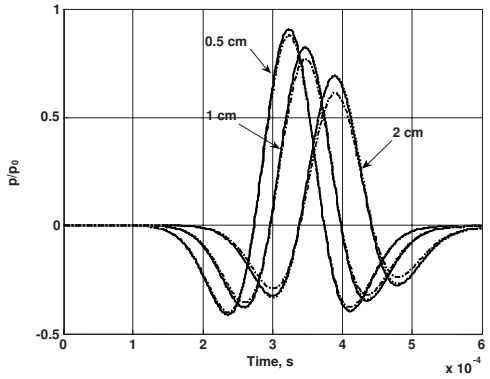


FIG. 2. Propagation of the short pulse in the packing of spheres. Solid line — equivalent fluid model (8) and (9), doted line—high frequency approximation (15) and (16), dash-dotted line—relaxational model (Ref. 8) with relaxation times defined by Eq. (18).

$\exp(-(t-t')/\tau_2)$ varies much slower than $\partial_{t'}V(t')/\sqrt{t-t'}$. Consequently, in the transient regime, the momentum conservation equation takes the following form:

$$\rho_0\alpha_\infty\partial_t V + \frac{\rho_0\alpha_\infty}{\tau_1}\sqrt{\frac{\tau_2}{\pi}}\int_{-\infty}^{t}\frac{\partial_{t'}V(t')}{\sqrt{t-t'}}dt' = -\partial_x P. \tag{15}$$

The first integral $I_1(t)$ of the continuity equation (9) can be simplified as

$$I_1(t) \approx \frac{\gamma-1}{T_1}\sqrt{\frac{T_2}{\pi}}\int_{-\infty}^{t}\frac{\partial_{t'}P(t')}{\sqrt{t-t'}}dt'.$$

The sum of other two integrals is approximated as:

$$I_2(t) + I_3(t) \approx -\frac{(\gamma-1)T_2}{T_1^2}\int_{-\infty}^{t}\partial_{t'}P(t')dt'.$$

In the transient regime it is much smaller than integral $I_1(t)$ and the continuity equation is reduced to the following:

$$\partial_t P + \frac{\gamma-1}{T_1}\sqrt{\frac{T_2}{\pi}}\int_{-\infty}^{t}\frac{\partial_{t'}P(t')}{\sqrt{t-t'}}dt' = -\rho_0 c^2\partial_x V. \tag{16}$$

It can easily be seen now that Eqs. (15) and (16) are exactly the same as Eqs. (25) and (26) of Ref. 5 formulated for the transient regime.

III. NUMERICAL RESULTS

Numerical simulations were performed in order to compare predictions given by different models. The numerical scheme adopted was the forward time centered space scheme.

The convolution integrals present in the equations were computed applying the procedure proposed elsewhere.[9] It is based on the assumption that if equations are discretized using a very short time step, then the functions of pressure and velocity in the convolution integrals can be considered constant during each step. This procedure considerably reduces the computational cost of the simulations.

It is assumed that the acoustic source produces a pulse defined by the following function:

$$f(t) = (1-\omega_0^2(t-t_0)^2)\exp\left(-\frac{\omega_0^2(t-t_0)^2}{2}\right).$$

known as a "Mexican hat."

The pulse is considered as "short" if its central frequency f_0 is at least five times higher than the critical frequency of the material $f_c = 1/2\pi\tau_2$ and "long" if its central frequency is at least one-fifth of the critical frequency. Between these limits the pulse is defined as being of "medium duration." Pulse parameters and are given in Table I.

The simulations have been performed for the packing of identical spherical particles with radius 0.5 mm and packing porosity equal to 0.36 (f_c=868.3 Hz); and gravel with maximum particle dimension 9.02 mm (f_c=10.6 Hz). Viscous parameters of the spherical packing (σ_0=27888 Pa s/m^2, α_∞=1.89, Λ=0.134 mm) have been obtained using the cell model,[12] while those of the gravel have been found in literature.[13] Thermal parameters were obtained using relationship valid for spherical packings.[14] In computations the samples have always been considered infinitely long. Pulses at different locations inside the material have been computed using time domain equivalent fluid model defined by Eqs. (8) and (9), low frequency/long pulse approximations (11) and

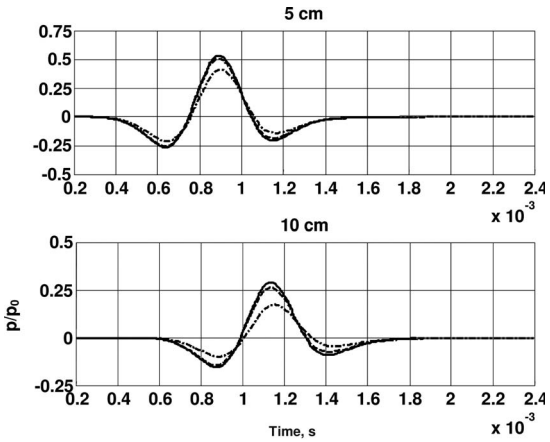


FIG. 3. Propagation of the medium duration pulse in the packing of spheres. Solid line—equivalent fluid model (8) and (9), dashed line—relaxational model (Ref. 8) with relaxation times given by Eq. (17), and dash-dotted line—relaxational model with relaxation times given by Eq. (18).

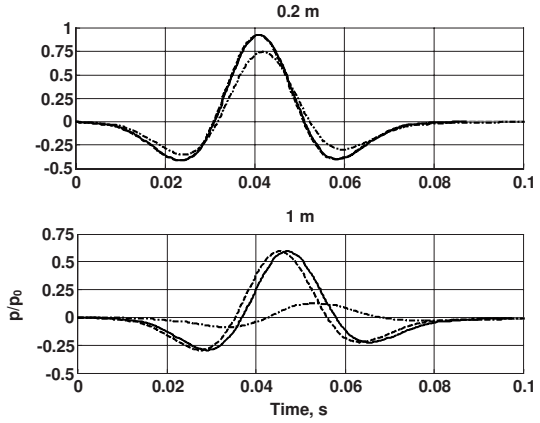


FIG. 4. Propagation of the medium duration pulse in gravel. Legend as in Fig. 3.

(14), high frequency/short pulse approximations (15) and (16), and relaxational model.[8] In the latter two sets of relaxation times have been used:

$$\tau_\nu = \frac{2\rho_0\alpha_\infty}{\phi\sigma_0} \text{ and } \tau_e = \frac{2N_{\mathrm{Pr}}k'\rho_0}{\phi\eta}, \tag{17}$$

which satisfy physically correct low frequency limits of complex density and complex compressibility functions,[1–3] and

$$\tau_\nu = \frac{\Lambda^2\rho_0}{4\eta} \text{ and } \tau_e = \frac{N_{\mathrm{Pr}}\Lambda'^2\rho_0}{4\eta}, \tag{18}$$

which satisfy those limits formulated for high frequencies.

The propagation of the long pulse in the packing of spheres is illustrated in Fig. 1 where its shape at three different distances from the source is shown. The equivalent fluid model predictions are very close to those of its low frequency limit and the relaxational model at all locations.

As shown in Fig. 2, the equivalent fluid model is close to its high frequency limit, whereas the relaxational model diverges slightly as the distance from the source increases.

For the medium duration pulse the comparisons between equivalent fluid model and the relaxational model predictions have been performed. In this range the values of the relaxation times have to be considered as adjustable parameters.[7] However, here we use two pairs defined by Eqs. (17) and (18) instead. It can be noted from Fig. 3, that in the packing of spheres the relaxational model with relaxation times defined by Eq. (17) remains close to the equivalent fluid model even when pulse central frequency is close to f_c. A higher discrepancy was found between the models for a pulse propagating in gravel (Fig. 4). This is due to the fact that the shape factor values for this material, $M=10.03$ and $M'=6.03$, are high and consequently pairs of the relaxation times (17) and (18) are more sharply tuned to satisfy low and high frequency limits. In spherical packing the shape factor values, $M=1.57$ and $M'=1.12$, are both close to unity.

The equations for the low frequency limit of the equivalent fluid model are obviously the fastest to solve, as they do not involve any convolutions. The amount of computational time required for the equivalent fluid model is comparable to other models.

IV. CONCLUSIONS

It has been shown that the equivalent fluid, model allows analytical formulation in the time domain. The derived set of equations can be applied to acoustical pulses without any limitations on their duration provided that wavelength exceeds pore/inclusion size. For the pulses with medium duration numerical solutions are relatively close to those of the time domain relaxational model for the materials with viscous and thermal shape factors close to unity. For materials with larger shape factor values, the two models give considerably different results.

ACKNOWLEDGMENTS

The work is supported by EPSRC Grant No. EP/E016529/1.

[1]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).
[2]Y. Champoux and J.-F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media," J. Appl. Phys. **70**, 1975–1979 (1991).
[3]D. Lafarge, P. Lemariner, J.-F. Allard, and V. Tarnow, "Dynamic compressibility of air in porous structures at audible frequencies," J. Acoust. Soc. Am. **102**, 1995–2006 (1997).
[4]S. R. Pride, F. D. Morgan, and A. F. Gangi, "Drag forces in porous-medium acoustics," Phys. Rev. B **47**, 4964–4978 (1993).
[5]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).
[6]Z. E. A. Fellah, M. Fellah, W. Lauriks, and C. Depollier, "Direct and inverse scattering of transient acoustic waves by a slab of rigid porous material," J. Acoust. Soc. Am. **113**, 61–72 (2003).
[7]D. K. Wilson, "Simple, relaxational models for the acoustical properties of porous media," Appl. Acoust. **50**, 171–188 (1997).
[8]D. K. Wilson, V. E. Ostashev, and S. L. Collier, "Time-domain equations for sound propagation in rigid frame porous media," J. Acoust. Soc. Am. **116**, 1889–1892 (2004).
[9]D. K. Wilson, S. L. Collier, V. E. Ostashev, D. F. Aldridge, N. P. Symons, and D. H. Marlin, "Time-domain modelling of the acoustic impedance of porous surfaces," Acta. Acust. Acust. **92**, 965–975 (2005).
[10]D. K. Wilson, V. E. Ostashev, S. L. Collier, N. P. Symons, D. F. Aldridge, and D. H. Marlin, "Time-domain calculations of sound interactions with outdoor ground surfaces," Appl. Acoust. **68**, 173–200 (2007).
[11]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[12]O. Umnova, K. Attenborough, and K. M. Li, "Cell model calculations of the dynamic drag parameters in packings of spheres," J. Acoust. Soc. Am. **107**, 3113–3119 (2000).
[13]O. Umnova, K. Attenbrough, H. Shin, and A. Cummings, "Response of multiple rigid porous layers to high levels of continues acoustic excitation," J. Acoust. Soc. Am. **116**, 703–712 (2004).
[14]O. Umnova, K. Attenborough, and K. M. Li, "A cell model for the acoustical properties of packings of spheres," Acta Acust. **87**, 226–235 (2001).

Prediction of the acoustic field due to a point source over a porous layer (L)

Jean F. Allard
Laboratoire d'Acoustique de l'Université du Maine, Avenue Olivier Messiaen, 72000 Le Mans, France

(Received 28 September 2008; revised 19 January 2009; accepted 22 January 2009)

Comparisons are performed between the results of a simple classic formulation for the acoustic field created by a point source above a porous layer and an exact description. A semi-infinite layer and layers of finite thickness are considered. It is shown that for layers of finite thickness and small flow resistivity the use of the classic formulation can lead to large errors close to grazing incidence.
 [DOI: 10.1121/1.3081500]

PACS number(s): 43.55.Ev, 43.20.El, 43.20.Ye [KVH] Pages: 1864–1867

I. INTRODUCTION

A description of sound propagation from a point source over an impedance plane with a specific impedance Z_s was obtained by Chien and Soroka.[1] Let R_1 be the distance from the image S′ of the source S to the microphone M, R_2 the distance from the source to the microphone, z_1 and z_2 the heights of the source and the microphone, r the radial distance from the source to the microphone, and θ_0 the angle of specular reflection. Like in Ref. 1 the angle of specular reflection is the angle between the normal to the plane and S′M. It cannot be initially defined as an angle of incidence because an angle of incidence is related to a plane wave and a point source can be considered as a superposition of plane waves in different directions. The total pressure p_z is the sum of the direct pressure and the reflected pressure p_r,

$$p_z = \frac{\exp(ik_0R_2)}{R_2} + p_r,$$

$$p_r = [V(\theta_0) + (1 - V(\theta_0))F(w)]\frac{\exp(ik_0R_1)}{R_1}, \tag{1}$$

$$F(w) = 1 + i\sqrt{\pi}w\exp(w^2)\mathrm{erfc}(-w), \tag{2}$$

$$w = i\left(\tfrac{1}{2}ik_0R_1\right)^{1/2}[\cos\theta_0 + Z_0/Z_s]. \tag{3}$$

In these equations k_0 is the wave number in the free air and V is the reflection coefficient given by

$$V(\theta) = \frac{Z_s - Z_0/\cos\theta}{Z_s + Z_0/\cos\theta}, \tag{4}$$

where Z_0 is the characteristic impedance of the free air and θ is the classical angle of incidence for a plane wave which is equal to θ_0 in Eqs. (1) and (3). This description is an approximation valid under several conditions,

$$k_0R_1 \gg 1, \tag{5}$$

$$\cos\theta_0 \ll 1,$$

$$Z_0/|Z_s| \ll 1. \tag{6}$$

A description similar to the one of Chien and Soroka[1] was obtained under the same conditions with the pole subtraction method by Brekhovskikh and Godin[2] [see Eq. (1.4.10) of Ref. 2]. Like in Ref. 1, the description of the reflected field is based on the presence of a pole of the reflection coefficient at an angle of incidence θ_p given by

$$\cos\theta_p = -Z_0/Z_s. \tag{7}$$

In the square brackets of Eq. (3) $\cos\theta_0 + Z_0/Z_s$ can be replaced by $\cos\theta_0 - \cos\theta_p$. In Refs. 1 and 2 an important parameter that characterizes the reflection of the spherical field close to grazing incidence is the numerical distance w that depends on the distance from θ_p to θ_0 via $\cos\theta_0 - \cos\theta_p$.

The impedance Z_s does not depend on the angle of incidence, whereas the surface impedance of a porous layer of small flow resistivity can noticeably depend on the angle of incidence. The formulation by Chien and Soroka[1] was modified to account for the extended reaction of porous layers by Nicolas *et al.*[3] and Li *et al.*[4] In the modified formulation the constant impedance Z_s is replaced in Eqs. (3) and (4) by the impedance $Z_s(\theta_0)$ of the layer at an angle of incidence equal to θ_0. The surface impedance at an angle of incidence θ for a porous layer of finite thickness having a motionless frame is given by

$$Z_s(\theta) = \frac{iZ_1}{\phi\cos\theta_1}\cot(k_1l\cos\theta_1), \tag{8}$$

which becomes for a semi-infinite layer $Z_s(\theta) = Z_1/(\phi\cos\theta_1)$. In these expressions ϕ is the porosity, k_1 is the wave number, Z_1 is the characteristic impedance for the waves which propagate in the saturating air in the porous medium, and θ_1 is the complex refraction angle given by

$$\sin\theta_1 = \frac{k_0}{k_1}\sin\theta. \tag{9}$$

There is no rigorous theoretical justification of the modified formulation. Moreover the initial formulation is valid under many restrictive conditions. An exact integral expression of the reflected spherical field which is always valid when the source and the microphone are located above the porous layer can be obtained with the Sommerfeld representation[2] of the sound field. The exact reflected field can be written as

TABLE I. Parameters for the porous media.

Medium	1	2	3
Porosity ϕ	0.98	0.98	0.3
Flow resistivity σ (N m^{-4} s)	10 000	30 000	2000
Thermal permeability k_0' (m^2)	1.84×10^{-19}	6.1×10^{-10}	9.2×10^{-9}
Viscous dimension Λ (μm)	120	70	700
Thermal dimension Λ' (μm)	360	210	2100
Tortuosity α_∞	1.1	1.1	2

$$p_r = i\int_0^\infty \xi d\xi \frac{V(\varphi)}{\mu} J_0(r\xi)\exp[i\mu(z_1+z_2)], \tag{10}$$

where $\mu=\sqrt{k_0^2-\xi^2}$, $\mathrm{Im}\,\mu\geqslant 0$, and φ is defined by $\cos\varphi=\mu/k_0$. The exact pressure field p_e is obtained by adding $\exp(ik_0R_2)/R_2$ to the exact reflected field. The singularity at $\mu=0$ can be removed by using μ instead of ξ as a variable of integration. For $\xi\gg 1$, $i\mu(z_1+z_2)$ is large and negative and beyond a finite value of ξ the contribution to the integral in Eq. (10) is negligible. The evaluation of the integral is easier for small r because the variations in $J_0(r\xi)$ as a function of ξ are slower, and for large z_1+z_2, because $\exp[i\mu(z_1+z_2)]$ decreases faster when μ is imaginary. The precision in the calculation of the integral can be evaluated for a given geometry by setting $V=1$ and comparing the result to $\exp(ik_0R_1)/R_1$. For instance, for a radial distances smaller than 10 m and z_1+z_2 larger than 4 cm, the trapezoidal method can be used to evaluate the integral, and for $r<0.5$ m the trapezoidal method can be used for $z_1+z_2\lessapprox 1$ cm. The modified formulation with p_r obtained by Eqs. (1)–(4) where Z_s is replaced by $Z_s(\theta_0)$ is currently used in long range sound propagation and to evaluate surface impedance and acoustical parameters of porous layers but as far as we know, there is no detailed evaluation of its reliability for thin layers with a small flow resistivity. Comparisons between predictions obtained with the exact integral expression for p_r and the modified formulation are performed for several media, foams with different low resistivities, and a porous pavement. A limitation concerning the use of the modified formulation is shown in the medium frequency range. Previous works are used to explain the good agreement between the modified formulation and the exact integral expression in the low and the high frequency range.

II. PRESSURE PREDICTIONS FOR NON-LOCALLY REACTING MEDIA

Three non-locally reacting media are considered, two materials with a porosity and a tortuosity close to 1, one similar to a foam with a low flow resistivity (medium 1) and the other similar to a foam with a larger flow resistivity (medium 2), and a material with a large tortuosity and a small porosity similar to a porous pavement (medium 3). The acoustical parameters of the media are given in Table I. The models used for the effective density and the bulk modulus of the air in the porous media are described in Ref. 5. The use of different models would lead to similar conclusions. The ratio p_z/p_e of the total pressure p_z predicted with the modified formulation [Eqs. (1)–(4)] with the constant impedance Z_s replaced by $Z_s(\theta_0)$ and the exact pressure p_e predicted with Eq. (10) is represented in Fig. 1 for medium 1 and medium 2. The thickness of the layer is $l=4$ cm. The geometry is defined by $z_1=z_2=0.025$ m and $r=1$ m. The angle of specular reflection $\theta_0=1.52$ rad is close to $\pi/2$. For both media, a good precision is obtained in the low and the high frequency range with the modified formulation. The precision is very poor in the medium frequency range for medium 1, but noticeably increases for medium 2 which has a higher flow resistivity because the dependence of the surface impedance on the angle of incidence decreases when the flow resistivity increases. The good precision in the low frequency range is due to the fact that when the thickness of the layer is small compared to the wavelength in the medium, $|k_1|l\ll 1$, a pole exists at θ_p close to $\pi/2$, the other poles are located far from the real sin θ axis and do not contribute to the reflected field. This pole is called the main pole.[6,7] The poles correspond to a 0 of the denominator of the reflection coefficient and $\cos\theta_p=-Z_0/Z_s(\theta_p)$. It has been shown[6,7] that the modified Chien and Soroka[1] formulation can be used to predict the reflected field with one change in the modified equation (3) where $Z_0/Z_s(\theta_0)$ must be replaced by $-\cos\theta_p=Z_0/Z_s(\theta_p)$. Then the change consists in replacing in the modified formulation $Z_0/Z_s(\theta_0)$ by $Z_0/Z_s(\theta_p)$ in the square brackets. In Fig. 2, $-Z_0/Z_s(\theta_0)$ and $-Z_0/Z_s(\theta_p)$ are represented as a function of frequency for the same layer of medium 1 and the same geometry. Both quantities are close to each other at low frequencies and the modification does not lead to noticeable variations in the predicted pressure. At high frequencies when $|k_1l|\geqslant 1$, the pole subtraction method cannot be used because there is no dominant pole close to $\pi/2$ but the passage path method can be used. The good agreement between p_z and p_e can be explained by using the evaluation of the reflected pressure with the passage path method. It has been shown by Brekhovskikh and Godin[2] that the reflected pressure p_r predicted with this method can be written under the condition of arbitrary dependence of the reflection coefficient on θ and for a product radial distance $\times$ wave number in air $k_0r\gg 1$ [see Eq. (1.2.9) of Ref. 2],

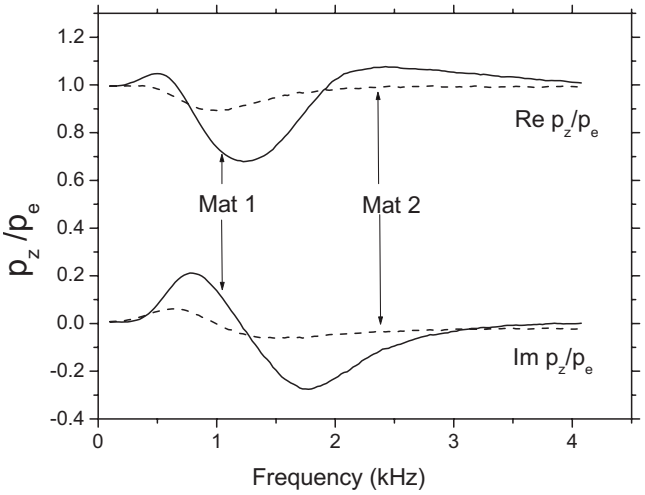


FIG. 1. Ratio p_z/p_e for layers of thickness $l=4$ cm of medium 1 — and medium 2 - - -, p_z pressure predicted with the modified formulation, p_e exact pressure. Geometry: $z_1=z_2=0.025$ m and $r=1$ m.

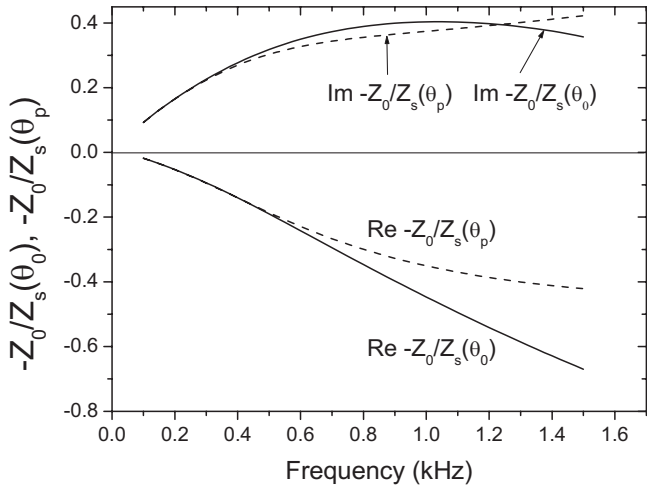


FIG. 2. Comparison between $\cos\theta_p = -Z_s(\theta_p)/Z_0$ for the main pole and $-Z_s(\theta_p)/Z_0$ for the same geometry and the same layer of medium 1 as in Fig. 1.

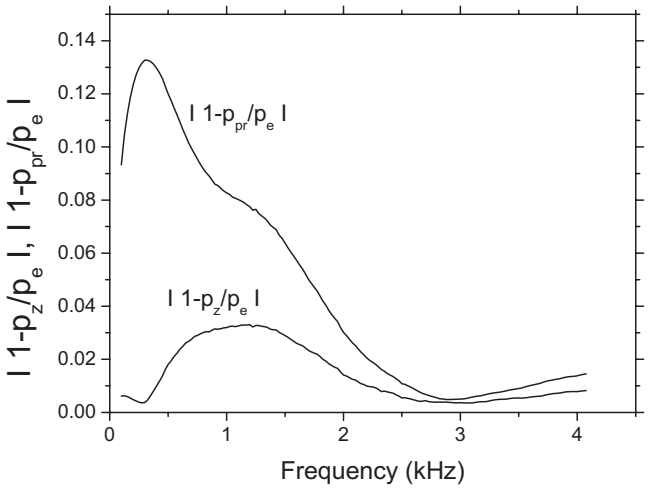


FIG. 3. $|1-p_z/p_e|$ compared to $|1-p_{pr}/p_e|$, p_{pr} pressure predicted with Eq. (14), p_z pressure predicted with the modified formulation, p_e exact pressure. Geometry: $z_1=0.025$ m, $z_2=0.6$ m, and $r=1$ m. Layer of thickness $l=4$ cm of medium 1.

$$p_r = \frac{\exp(ik_0,R_1)}{R_1}\left[V(\theta_0) - \frac{iN}{k_0R_1}\right],$$

$$N = \frac{1}{2}\left[(1-\sin^2\theta)\frac{\partial^2 V}{\partial(\sin\theta)^2} + \frac{1-2\sin^2\theta}{2\sin\theta}\frac{\partial V}{\partial\sin\theta}\right]_{\theta\to\theta_0}. \quad (11)$$

The coefficient N has been calculated by Allard *et al.*[8] for a porous layer of finite thickness [see Eqs. (A.10) and (A.11) of Ref. 8]. At the zero order approximation in θ_0, N is given by

$$N = -2\left(\frac{Z_s(\theta_0)}{Z_0}\right)^2. \quad (12)$$

Using in Eq. (1) of the modified formulation the asymptotic value $1/(2w^2)$ for $F(w)$ and replacing close to grazing incidence $(1-V)$ by -2, Eq. (1) gives the same expression for p_r as Eqs. (11) and (12),

$$p_r = \frac{\exp(ik_0R_1)}{R_1}\left[V(\theta_0) + \frac{2i(Z_s(\theta_0)/Z_0)^2}{k_0R_1}\right]. \quad (13)$$

For a smaller angle of incidence $|1-p_z/p_e|$ is presented in Fig. 3 as a function of frequency for a layer of thickness $l=4$ cm. The new geometry is defined by a radial distance $r=1$ m, $z_1=0.025$ m, and $z_2=0.6$ m. The contribution of the added term $(1-V(\theta_0))F(w)$ to the plane wave reflection coefficient $V(\theta_0)$ is now less important. Let p_{pr} be the predicted pressure obtained by removing the added term

$$p_{pr} = \frac{\exp(ik_0R_2)}{R_2} + V(\theta_0)\frac{\exp(ik_0R_1)}{R_1}. \quad (14)$$

The quantity $|1-p_{pr}/p_e|$ is also presented in Fig. 3. A good approximation for the exact pressure is obtained with p_{pr} but in the low frequency range p_z is much closer to p_e. One reason is the fact that at sufficiently low frequencies the porous medium tends to become locally reacting because $|k_1/k_0|$ increases when frequency decreases. Then the replacement of $Z_0/Z_s(\theta_0)$ by $Z_0/Z_s(\theta_p)$ in the square brackets of Eq. (3) has a negligible effect because $Z_s(\theta_0)$ and $Z_s(\theta_p)$ are close to each other even if $\cos\theta_p$ and $\cos\theta_0$ are noticeably different. For instance, $\cos\theta_0=0.614$ and at 0.5 kHz $\cos\theta_p=-Z_0/Z_s(\theta_p)=-0.186+i0.303$, and $-Z_0/Z_s(\theta_0)=-0.157+i0.350$. The second reason is that the condition $\cos\theta_0 \ll 1$ in Eq. (6) is not drastic.

Pressure variation for two heights of the receiver as a function of frequency above a porous layer is used to evaluate the surface impedance or the acoustic parameters of the medium. For a layer of thickness $l=4$ cm of medium 3 the ratio of the pressure p' for $z_1=0.025$ m, $z_2=0.6$ m, and $r=1$ m over p for $z_1=z_2=0.025$ m and $r=1$ m is presented in Fig. 4. The true ratio p'_e/p_e is compared with the predicted ratio p'_z/p_z obtained with the modified formulation. Large differences between both ratios appear in the medium frequency range, which can lead to noticeable errors in the evaluation of the acoustical parameters for this medium of low flow resistivity.

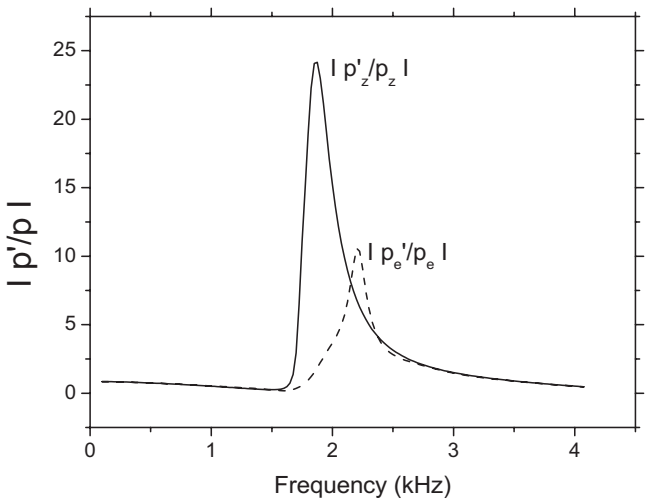


FIG. 4. Modulus of the exact ratio of pressures p'_e at $z_2=0.6$ m and p_e at $z_2=0.025$ m - - -, modulus of the ratio of the predicted pressures with the modified formulation p'_z and p_z —. For both heights of the receiver, $z_1=0.025$ m and $r=1$ m. Layer of thickness $l=4$ cm of medium 3.

TABLE II. Max $|1-p_z/p_e|$ as function of thickness for a layer of medium 1, $z_1=z_2=0.025$ m and $r=1$ m.

Thickness (cm)	2	4	8	∞
Max $\|1-p_z/p_e\|$	0.43	0.33	0.22	0.11
Frequency (kHz)	3.1	1.35	0.65	0.75

A first conclusion after these modelizations is that for thin sound absorbing porous materials having a flow resistivity smaller than 30 000 N m^{-4} s the modified formulation can lead to significant errors in the prediction of the pressure close to grazing incidence. In a second step the dependence on the thickness of the reliability of the modified formulation has been precised in the following way. For medium 1 the maximum of $|1-p_z/p_e|$ as a function of frequency is given in Table II for different thicknesses from 2 cm to infinity for $z_1=z_2=0.025$ m and $r=1$ m. The maximum error increases when the thickness decreases and the maximum reaches the high frequency range. The modified formulation is more reliable for semi-infinite layers. This is due to the fact that like for the case of a plane impedance there is only one pole of the reflection coefficient located at θ_p given by

$$\cos\theta_p=\left(\frac{n^2-1}{(\rho_1/\rho_0\phi)^2-1}\right)^{1/2},$$

$$\mathrm{Re}\cos\theta_p<0, \tag{15}$$

where ρ_0 is the density of air, ρ_1 is the effective density in the porous medium, and $n=k_1/k_0$. Moreover the variations in the surface impedance as a function of the angle of incidence are smaller for a semi-infinite layer than for a layer of finite thickness made of the same porous medium. As indicated previously, the accuracy of the modified formulation increases when the dependence decreases. The expression of the reflected spherical field obtained by Brekhovskikh and Godin[2] at the interface between two semi-infinite fluids [Eq. (1-4-4)] can be used for a semi-infinite porous layer, the density ratio m being replaced by $\rho_1/(\rho_0\phi)$. It can be shown that if the conditions (6) are fulfilled, Eq. (1-4-4) of Ref. 2 leads to the same system of equations as for a thin porous layer, the surface impedance being now given by $Z_s(\theta)=Z_1/(\phi\cos\theta_1)$ and the pole location by Eq. (15).

III. CONCLUSION

Predictions of the reflected spherical point source field obtained from the description by Chien and Soroka[1] modified to account for the non-locally reacting behavior of porous layers, and obtained from an exact model, have been compared. The modified formulation can provide reliable predictions for media with a sufficiently large flow resistivity, but the predictions can present large errors in the medium frequency range for thin layers made of media with a small flow resistivity. For these materials, for small radial distances from the source to the receiver, the modified description should be replaced by the exact description obtained from the Sommerfeld representation.

[1] C. F. Chien and W. W. Soroka, "Sound propagation along an impedance plane," J. Sound Vib. **43**, 9–20 (1975).
[2] L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media II, Point Source and Bounded Beams*, Springer Series on Wave Phenomena (Springer, New York, 1992).
[3] J. Nicolas, J. L. Berry, and G. A. Daigle, "Propagation of sound above a layer of snow," J. Acoust. Soc. Am. **77**, 67–73 (1985).
[4] K. M. Li, T. Waters-Fuller, and K. Attenborough, "Sound propagation from a point source over extended-reaction ground," J. Acoust. Soc. Am. **104**, 679–685 (1998).
[5] 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).
[6] J. F. Allard, M. Henry, V. Gareton, G. Jansens, and W. Lauriks, "Impedance measurement around grazing incidence for nonlocally reacting thin porous layers," J. Acoust. Soc. Am. **113**, 1210–1215 (2003).
[7] W. Lauriks, L. Kelders, and J. F. Allard, "Poles and zeros of the reflection coefficient of a porous layer having a motionless frame in contact with air," Wave Motion **28**, 59–67 (1998).
[8] J. F. Allard, G. Jansens, and W. Lauriks, "Reflection of spherical waves by non-locally reacting porous media," Wave Motion **36**, 143–155 (2002).

A filtered convolution method for the computation of acoustic wave fields in very large spatiotemporal domains

Martin D. Verweij[a] and Jacob Huijssen
Laboratory of Electromagnetic Research, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

(Received 22 July 2008; revised 8 January 2009; accepted 12 January 2009)

The full-wave computation of transient acoustic fields with sizes in the order of $100\times100\times100$ wavelengths by 100 periods requires a numerical method that is extremely efficient in terms of storage and computation. Iterative integral equation methods offer a good performance on these points, provided that the recurring spatiotemporal convolutions are computed with a coarse sampling and relatively few computational operations. This paper describes a method for the numerical evaluation of very large-scale, four-dimensional convolutions that employs a fast Fourier transformation and that uses a sampling rate close to or at the limit of two points per wavelength and per period. To achieve this, the functions involved are systematically filtered, windowed, and zero-padded with respect to all relevant coordinates prior to sampling. The method is developed in the context of the Neumann iterative solution of the acoustic contrast source problem for an inhomogeneous medium. The implementation of the method on a parallel computer is discussed. The obtained numerical results have a relative root mean square error of a few percent when sampling at two points per wavelength and per period. Further, the results prove that the method enables the computation of transient fields in the order of the indicated size.
 [DOI: 10.1121/1.3077220]

PACS number(s): 43.20.Px, 43.20.El, 43.20.Bi, 43.60.Gk [TDM] Pages: 1868–1878

I. INTRODUCTION

The computation of transient acoustic wave fields over very large spatiotemporal domains with inhomogeneous, dispersive, and nonlinear media plays an increasingly important role in the development and operation of novel acoustical applications and devices. For instance, consider the realization of an inverse scattering method for the accurate characterization of biomedical tissue using pulsed ultrasound data. The basis of this inverse method will be a full-wave, forward method that is repeatedly invoked for the computation of a transient acoustic pressure field in the inhomogeneous medium under reconstruction. In case of human organs and ultrasound frequencies, the forward computations may easily involve a domain that measures 100 wavelengths in three spatial directions by 100 periods in time. These forward computations provide a typical example of the kind and size of problems that have motivated the research reported in this paper.

Even with today's parallel computers, the numerical solution of full-wave, forward problems of the abovementioned size is only practically feasible if the applied numerical method is extremely efficient in terms of storage and computation. The largest classes of methods for the numerical solution of forward wave problems are finite difference methods, finite element methods, and integral equation methods. As is motivated below, in this paper iterative integral equation methods are selected for dealing with very large-scale, full-wave, forward problems. A variety of iterative methods exist for solving integral equations,[1–4] e.g., the Neumann iterative solution, over-relaxation methods, and conjugate gradient methods. All these methods aim to provide a sequence of increasingly accurate, global approximations to the wave field, and use convolutions over three-dimensional space and time to obtain these. In practice, the convolutions must be performed numerically, which requires the discretization of the computational domain with respect to its space and time coordinates. This paper describes a method for the numerical evaluation of convolutions over very large spatiotemporal domains. The discussion is limited to equidistant sampling and rectangular grids, and the symbol D_φ is used to indicate the number of grid points per wavelength and per period of a particular field component with temporal (angular) frequency φ. In this respect, the most relevant quantity is D_Φ, where Φ is the chosen maximum temporal angular frequency of interest.

With regard to storage, iterative integral equation methods require that one or more scalar field quantities should be available over the full spatiotemporal grid for further iteration. Since the total amount of grid points in the computational domain is proportional to D_Φ^4, it is paramount to make D_Φ as small as possible to avoid excessive memory requirements. However, the minimum value of D_Φ is dictated by the desired accuracy of the computations. The most memory-demanding operations in an iterative integral equation method are recurring convolutions of functions that depend on the stored iterates and kernel functions that are known in advance, e.g., the Green's function of a background medium. These convolutions involve the entire computational domain. As a consequence of the Nyquist–Shannon sampling theorem,[5] the sampling may in principle be as coarse as $D_\Phi=2$ to yield exact convolutional results at the grid points,

[a]Electronic mail: m.d.verweij@tudelft.nl

 0001-4966/2009/125(4)/1868/11/$25.00

provided that components with an angular frequency above Φ are absent. For the typical example presented above, sampling at the limit $D_\Phi=2$ results in a grid of 1.6×10^9 points, and a parallel implementation of the chosen method demands a memory size in the order of 100 Gbytes. With modern, high-performance computing facilities, this is very well feasible.

From a computational point of view, the numerical convolutions may benefit from using the Fast Fourier Transformation (FFT).[6] For a convolution involving N sample points, this requires only order $N\log(N)$ operations instead of order N^2 operations for a straightforward quadrature rule. In view of this and the fact that N can be kept low by sampling close to or at the limit $D_\Phi=2$, it may be concluded that iterative integral equation methods combine a minimal storage requirement with a strongly reduced computational effort.

For comparison it is noted that standard finite difference or finite element methods use relatively simple, local functions to approximate the wave field and need at least a grid with $D_\Phi=10$ to yield accurate results.[7,8] As a consequence, for the same computational domain these methods need much more storage than an iterative integral equation method. This becomes evident when realizing that the typical example mentioned above now requires 10^{12} grid points.

Although many authors[9–12] employ FFT's for the numerical evaluation of convolutions, this is rarely accompanied by a minimization of the number of grid points. Presumably, this has to do with the fact that a coarse sampling requires special preparations, and for most applications storage is not so much of a bottleneck as computation. Since this paper deals with very large-scale wave problems, the sampling issue is at least as relevant as the application of FFT's. Therefore, this paper focuses on the development of a method for the numerical evaluation of very large-scale, four-dimensional convolutions that is based on FFT's and that uses a sampling rate close to or at the limit $D_\Phi=2$. Since filtering forms the basis of achieving the coarse sampling, this method is referred to as the filtered convolution method. The filtered convolution method is generic because it is applicable in all numerical methods that employ convolutions.

The approach to obtain a coarse sampling is to select a temporal angular cutoff frequency $\Omega\geq\Phi$ and a corresponding spatial angular cutoff frequency K, and to apply ideal filters to remove all higher temporal and spatial frequencies from the quantities to be sampled. This enables temporal and spatial samplings with $D_\Omega=2$, without the occurrence of aliasing. In addition, ideal windows are used to truncate the domain of the quantities, and zero-padding is employed. These steps are necessary to avoid wraparound due to the circular nature of the FFT-based numerical convolution.[6] The use of FFT's implies that the sampling occurs in both the original $\boldsymbol{x},t$ domain and in the transformed $\boldsymbol{k},\omega$ domain, and therefore the filtering and the windowing must in principle be performed in both domains. However, ideal filtering and ideal windowing are dual operations under the Fourier transformation, so a particular operation in one domain is automatically performed when its counterpart is performed in the other domain.

Central in this paper is a systematic explanation of the method that leads to a memory efficient and computationally fast numerical convolution. To describe this in a context, the contrast source formulation of the forward acoustic wave problem is introduced, and the Neumann iterative solution is employed to solve it. The authors are aware that other iterative integral equation methods may be more suitable for solving the problem at hand. However, selection of the best iterative method is not the issue of this paper, and the Neumann iterative solution has been chosen because it provides the most straightforward application of the filtered convolution method.

II. PROBLEM FORMULATION

A. Contrast source problem

The filtered convolution method is developed in the context of solving the system of first-order acoustic equations:

$$\boldsymbol{\nabla}p+\rho_0\partial_t\boldsymbol{v}=\boldsymbol{f}_p+\boldsymbol{f}_c(p,\boldsymbol{v}), \tag{1}$$

$$\boldsymbol{\nabla}\cdot\boldsymbol{v}+\kappa_0\partial_t p=q_p+q_c(p,\boldsymbol{v}), \tag{2}$$

in which p is the acoustic pressure and $\boldsymbol{v}$ is the particle velocity. This so-called contrast source formulation[13,14] is particularly useful for problems involving a medium with a localized inhomogeneous, anisotropic, dispersive, or nonlinear behavior. Outside the localized region, the medium behaves like a homogeneous, isotropic, dispersionless, and linear background medium with a mass density ρ_0 and a compressibility κ_0. The difference between the mass density of the actual medium and the background medium is accounted for by the contrast source density of force $\boldsymbol{f}_c(p,\boldsymbol{v})$, and the difference between the actual compressibility and the background compressibility is covered by the contrast source density of injection rate $q_c(p,\boldsymbol{v})$. The primary source that generates the acoustic wave field is represented by the volume density of force $\boldsymbol{f}_p$ and the volume density of injection rate q_p.

In many cases, combination of Eqs. (1) and (2) leads to a complete elimination of $\boldsymbol{v}$ and results in an acoustic wave equation

$$c_0^{-2}\partial_t^2 p-\boldsymbol{\nabla}^2 p=S_p+S_c(p). \tag{3}$$

Here, $c_0=(\rho_0\kappa_0)^{-1/2}$ is the wave speed of the background medium, and $S_p=\rho_0\partial_t q_p-\boldsymbol{\nabla}\cdot\boldsymbol{f}_p$ represents the primary source. The contrast source $S_c(p)$ now accounts for all differences between the actual medium and the background medium. Examples of contrast sources are

$$S_{c,\mathrm{inh}}(p)=[c_0^{-2}-c^{-2}(\boldsymbol{x})]\partial_t^2 p-\boldsymbol{\nabla}\ln[\rho(\boldsymbol{x})]\cdot\boldsymbol{\nabla}p, \tag{4}$$

$$S_{c,\mathrm{dis}}(p)=[c_0^{-2}\delta(t)-\rho(t)*_t\kappa(t)]*_t\partial_t^2 p, \tag{5}$$

$$S_{c,\mathrm{nl}}(p)=\frac{\beta}{\rho_0 c_0^4}\partial_t^2 p^2, \tag{6}$$

which, respectively, apply to an inhomogeneous medium with spatially varying $c(\boldsymbol{x})$ and $\rho(\boldsymbol{x})$, an attenuative and dispersive medium with relaxation functions $\rho(t)$ and $\kappa(t)$, and a

nonlinear medium with a coefficient of nonlinearity β. The symbol $*_t$ denotes a temporal convolution.

For the background medium, the Green's functions of the first-order acoustic equations and the acoustic wave equation are available in closed form.[15] This enables a solution of the corresponding contrast source problems by an iterative integral equation method, as shown next.

B. Neumann iterative solution

Knowing the Green's function G of the wave equation for the background medium, Eq. (3) may be recast into the integral equation[16,17]

$$p = G*_{\boldsymbol{x},t}[S_p + S_c(p)] = \mathcal{L}[S_p + S_c(p)]. \tag{7}$$

Here, the symbol $*_{\boldsymbol{x},t}$ denotes a convolution over the entire spatiotemporal computational domain, and it is assumed that the support of the primary and contrast source is located entirely within this domain. The convolution with the Green's function forms an integral operator, which is indicated by the symbol $\mathcal{L}$. Now it is supposed that the influence of the contrast source on the total field is relatively weak, i.e., that $\delta p = \mathcal{L}[S_c(p)]$ is small in comparison to $p^{(0)} = \mathcal{L}[S_p]$. Then successive approximations $p^{(j)}$ $(j=0,1,2,\ldots)$ to p may be generated by the Neumann iterative solution[3,4,18]

$$p^{(0)} = \mathcal{L}[S^{(0)}], \quad S^{(0)} = S_p, \tag{8}$$

$$p^{(j)} = p^{(0)} + \mathcal{L}[S^{(j)}], \quad S^{(j)} = S_c(p^{(j-1)}), \quad j \geq 1. \tag{9}$$

To get $\mathcal{L}[S^{(j)}] = \mathcal{L}[S^{(j)}](\boldsymbol{x},t) = \delta p^{(j)}(\boldsymbol{x},t)$, it is necessary to evaluate the four-dimensional convolution integral

$$\mathcal{L}[S^{(j)}](\boldsymbol{x},t) = \int_{\mathcal{D}_T^{(j)}} \int_{\mathcal{D}_X^{(j)}} G(\boldsymbol{x}-\boldsymbol{x}', t-t') S^{(j)}(\boldsymbol{x}',t') d\boldsymbol{x}' dt'. \tag{10}$$

For $j=0$ the integration is over the four-dimensional support $\mathcal{D}_T^{(0)} \times \mathcal{D}_X^{(0)}$ of the primary source $S^{(0)}$, and for $j \geq 1$ the integration is over the four-dimensional support $\mathcal{D}_T^{(j)} \times \mathcal{D}_X^{(j)}$ of the approximate contrast source $S^{(j)}$.

Even in the simplest cases, the convolution integral in Eq. (10) needs to be discretized and numerically evaluated. As has been discussed in Sec. I, for very large-scale problems it is vital to use a sampling that is as coarse as possible, preferably down to the limit $D_\Phi = 2$. In Secs. III and IV, it will be explained how this limit may be reached in case of a one-dimensional (temporal) convolution integral and in case of the four-dimensional (spatiotemporal) convolution integral in Eq. (10).

III. ACHIEVING $D_\Phi = 2$ IN CASE OF A ONE-DIMENSIONAL CONVOLUTION INTEGRAL

A. Basic Fourier theorems

First, it is convenient to recall two basic properties of an arbitrary function $f(t)$ and its Fourier transform

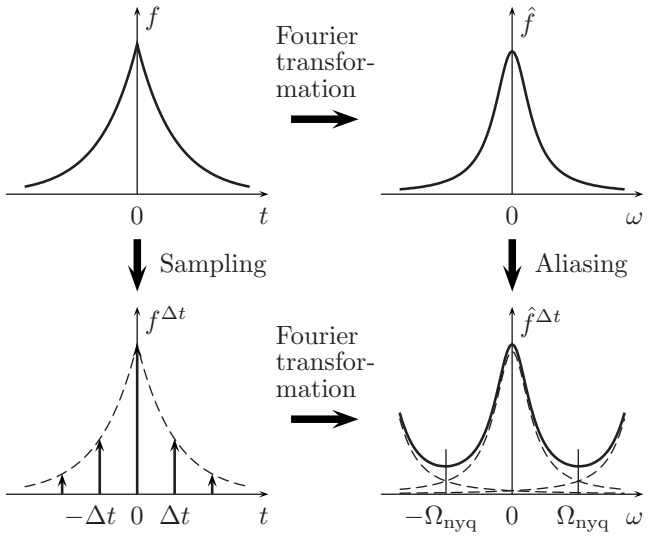


FIG. 1. Illustration of aliasing in the Fourier domain due to the sampling of a function in the original domain.

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-j\omega t) dt. \tag{11}$$

The first property involves the process of equidistant sampling of $f(t)$. If a sampled version of $f(t)$ is represented by

$$f^{\Delta t}(t) = \sum_{n=-\infty}^{\infty} f(n\Delta t)\, \delta\left(\frac{t-n\Delta t}{\Delta t}\right), \tag{12}$$

where Δt is the sampling interval, then the Fourier transform of this function is

$$\hat{f}^{\Delta t}(\omega) = \sum_{m=-\infty}^{\infty} \hat{f}(\omega + m\Omega_{\mathrm{per}}), \tag{13}$$

which is periodic with $\Omega_{\mathrm{per}} = 2\pi/\Delta t$. In the dual case, sampling of $\hat{f}(\omega)$ with a sampling interval $\Delta\omega$ results in a function $f^{\Delta\omega}(t)$ that is periodic with $T_{\mathrm{per}} = 2\pi/\Delta\omega$. If $\hat{f}(\omega)$ is nonzero for $|\omega| > \Omega_{\mathrm{per}}/2$, the terms in Eq. (13) overlap and $\hat{f}^{\Delta t}(\omega) \neq \hat{f}(\omega)$ for $|\omega| < \Omega_{\mathrm{per}}/2$. This is the cause of aliasing and is illustrated in Fig. 1. The quantity $\Omega_{\mathrm{per}}/2 = \pi/\Delta t = \Omega_{\mathrm{nyq}}$ is the angular Nyquist frequency.

The second property involves the restriction of the support of $f(t)$ to an interval $[-T,T]$ according to

$$f^T(t) = f(t)[H(t+T) - H(t-T)], \tag{14}$$

where $H(t)$ is the Heaviside step function and $2T$ is the size of the applied rectangular window. The Fourier transform of this function is

$$\hat{f}^T(\omega) = \frac{T}{\pi} \hat{f}(\omega) *_\omega \mathrm{sinc}(T\omega), \tag{15}$$

in which $\mathrm{sinc}(x) = \sin(x)/x$ is the unnormalized sinc function. In the dual case, restriction of the support of $\hat{f}(\omega)$ to an interval $[-\Omega,\Omega]$ results in a function $f^\Omega(t)$ that is given by

$$f^{\Omega}(t) = \frac{\Omega}{\pi} f(t) *_t \mathrm{sinc}(\Omega t). \tag{16}$$

A finite support in the original domain always leads to an infinite support in the Fourier domain, and vice versa. In this paper, the operations to get $f^T(t)$ or $\hat{f}^T(\omega)$ are referred to as *windowing* and the operations to obtain $f^{\Omega}(t)$ or $\hat{f}^{\Omega}(\omega)$ are referred to as *filtering*.

B. Straightforward discretization of a convolution integral

Second, it is relevant to consider the numerical evaluation of a one-dimensional convolution integral

$$F(t) = G(t) *_t S(t) = \int_{-\infty}^{\infty} G(t-t')S(t')dt'. \tag{17}$$

A straightforward approach is to use the left Riemann sum with a sampling interval Δt and to approximate the convolution integral by the convolution sum[19]

$$F_n = G_n *_n S_n = \Delta t \sum_m G_{n-m} S_m. \tag{18}$$

Here, $G_n = G(n\Delta t)$ and $S_n = S(n\Delta t)$ exactly, but F_n is only an approximation of $F(t)$ at the collocation points $n\Delta t$. The exclusive use of the function values at the collocation points implies that the functions have been sampled. To keep the number of samples finite, both $n-m$ and m must be limited to a finite number, which means the windowing of $G(t)$, $S(t)$, and, consequently, $F(t)$. When the given support of S_n is $[0, N-1]$ and the desired interval of interest of F_n is also $[0, N-1]$, then the support of G_n may be restricted to $[-N+1, N-1]$ without consequences for F_n on the desired interval. For efficiency, the convolution sum is usually evaluated by applying FFT's.[20] Since this implies a circular convolution, wraparound must be avoided *a priori* by supplementing S_n with zero-valued samples S_{-N+1} to S_{-1} (zero-padding).[6] The computational effort of the FFT's thus employed is of order $2N \log(2N)$.[6,19] The remaining approximation error $F_n - F(n\Delta t)$ is due to the sampling and windowing operations. For a typical pulse $S(t)$ with a central angular frequency Ω_0, an accurate evaluation of the components of $F(t)$ up until an angular frequency of interest $\Phi = 2\Omega_0$ may easily require a sampling with $D_\Phi = 10$ or more when no prior measures are taken. This may result in a prohibitively large amount of grid points, especially for convolutions in more dimensions.

C. Coarse discretization of a convolution integral

Now the stage has been set for a discussion of the proposed method for reducing D_Φ while keeping the approximation error under control. The underlying assumption is that a maximum angular frequency of interest Φ can be chosen in such a way that all components of interest in $F(t)$ have an angular frequency $|\omega| \leq \Phi$. This implies that limitation of the spectra of $G(t)$ and $S(t)$ by filtering these functions according to Eq. (16) with an angular cutoff frequency $\Omega \geq \Phi$ only removes insignificant components of $F(t)$. By subsequently taking a sampling interval $\Delta t = \pi/\Omega$, which implies $D_\Phi = 2\pi/\Phi\Delta t = 2\Omega/\Phi$, the situation $\Omega = \Omega_{\mathrm{nyq}}$ is enforced and aliasing is avoided. This means that F_n is obtained as the exact value of $F^{\Omega}(n\Delta t)$, and the approximation error is entirely due to the difference between $F(n\Delta t)$ and $F^{\Omega}(n\Delta t)$, which manifests itself only in the part of the spectrum that is of no interest as long as $D_\Phi \geq 2$. With the proposed method, preference is thus given to discarding the contributions from the frequencies $|\omega| > \Omega$ over making an aliasing error in the frequency range $|\omega| \leq \Omega$.

So far, the approach seems ideal. However, computational reality demands a finite number of samples, and this enforces the windowing of the functions involved. Assuming that the domain of interest ranges from $t=0$ to T, a window with a size $2T$ as given in Eq. (14) is applied to both $G^{\Omega}(t)$ and $S^{\Omega}(t)$. This requires the prior formal extension of the support of $S(t)$ by defining $S(t)=0$ for $-T \leq t < 0$. The given extension corresponds to the zero-padding that is required later on. The application of the window to $G^{\Omega}(t)$ and $S^{\Omega}(t)$, in fact, re-introduces an infinite support of $\hat{G}^{\Omega,T}(\omega)$ and $\hat{S}^{\Omega,T}(\omega)$—the order of the superscripts indicating the order of corresponding operations—that will lead to aliasing. In practice, the associated error is only a fraction of the aliasing error that would occur with a direct sampling of $G(t)$ and $S(t)$.

To summarize, the filtered convolution method for the coarse discretization of a one-dimensional convolution integral involves the following:

(1) Subjecting $G(t)$ and $S(t)$ to an ideal low-pass filter with an angular cutoff frequency $\Omega \geq \Phi$, and to a time window with a size $2T = (2N-1)\pi/\Omega$.
(2) Sampling $G^{\Omega,T}(t)$ and $S^{\Omega,T}(t)$ at $2N$ points with a sampling interval $\Delta t = \pi/\Omega$.
(3) Obtaining the Fourier transforms of G_n and the zero-padded S_n using a $2N$-point FFT.
(4) Multiplying the Fourier transforms and return to the original domain using a $2N$-point inverse FFT.

This yields $F^{\Omega,T}(n\Delta t)$ as an approximation for $F(n\Delta t)$.

Alternatively, if both $\hat{G}(\omega)$ and $\hat{S}(\omega)$ are known, the dual operations may be performed in the transform domain. In this case it is convenient to first perform the windowing. Now the application of the filter to $\hat{G}^T(\omega)$ and $\hat{S}^T(\omega)$ re-introduces an infinite support of $G^{T,\Omega}(t)$ and $S^{T,\Omega}(t)$ that will give rise to time-domain aliasing. Usually, the corresponding error is only a fraction of the time-domain aliasing error that would occur with a direct sampling of $\hat{G}(\omega)$ and $\hat{S}(\omega)$.

IV. ACHIEVING $D_\Phi = 2$ IN CASE OF A FOUR-DIMENSIONAL CONVOLUTION INTEGRAL

A. Extension of the one-dimensional method

For the numerical evaluation of the four-dimensional convolution integral in Eq. (10), the method from Sec. III is applied to all four coordinates of both $G(\boldsymbol{x},t)$ and $S^{(j)}(\boldsymbol{x},t)$. The filtering and windowing operations for the spatial dimensions are denoted by the subscripts K and X, respectively.

The temporal angular cutoff frequency Ω is taken as the prime parameter for both the temporal and the spatial discretizations. As explained in Sec. III C, this quantity directly determines the temporal sampling interval $\Delta t=\pi/\Omega$. The key to the spatial sampling is the decomposition of a wave in three-dimensional space into planar wave components with real angular wavevectors $\boldsymbol{k}$, as provided by the spatial Fourier representation. For a given Ω it is now assumed sufficient to only retain wave components with $|\boldsymbol{k}|=(k_x^2+k_y^2+k_z^2)^{1/2}\leq K=\Omega/c_0$. This choice for the spatial angular cutoff frequency K is justified if the smallest details in the configuration measure at least several wavelengths at Ω. In that case the width of the spatial spectrum of the contrast source mainly depends on that of the field, and since all solutions of the source-free wave equation for the background medium have $|\boldsymbol{k}|=\omega/c_0$, the choice $K=\Omega/c_0$ seems appropriate. The intended use of FFT's implies that k_x, k_y, and k_z stay real. Consequently, in each spatial dimension the maximum possible wavenumber is K and the resulting spatial sampling interval is $\Delta x=\pi/K=\pi c_0/\Omega$. To prevent spatial aliasing with the given spatial sampling, it is *necessary* to ensure that $|k_x|\leq K$, $|k_y|\leq K$, and $|k_z|\leq K$, i.e., to apply an ideal cube-shaped filter in the spatial Fourier domain. However, as stated above, it is assumed *sufficient* to ensure that $|\boldsymbol{k}|\leq K$, i.e., to apply an ideal spherical filter in the spatial Fourier domain.

B. Spatiotemporal filtering and windowing of the four-dimensional Green's function

The four-dimensional, background Green's function occurring in Eq. (10) is[15–17]

$$G(\boldsymbol{x},t)=\frac{\delta(t-|\boldsymbol{x}|/c_0)}{4\pi|\boldsymbol{x}|}. \tag{19}$$

Its temporal Fourier domain equivalent is

$$\hat{G}(\boldsymbol{x},\omega)=\frac{\exp(-jk|\boldsymbol{x}|)}{4\pi|\boldsymbol{x}|}, \tag{20}$$

with $k=\omega/c_0$, and its spatiotemporal Fourier domain counterpart is

$$\tilde{G}(\boldsymbol{k},\omega)=\frac{1}{|\boldsymbol{k}|^2-k^2}. \tag{21}$$

It is convenient to first perform the spatial filtering in the spatiotemporal Fourier domain. Here, the spherical symmetry of $\tilde{G}(\boldsymbol{k},\omega)$ in $\boldsymbol{k}=0$ is used by applying a spherical filter with an angular cutoff frequency K in the spatial Fourier domain, giving

$$\tilde{G}_K(\boldsymbol{k},\omega)=\tilde{G}(\boldsymbol{k},\omega)H(K-|\boldsymbol{k}|). \tag{22}$$

The temporal Fourier domain equivalent of this equation has been derived by Van der Veeken and Blok[21] as

$$\begin{aligned}\hat{G}_K(\boldsymbol{x},\omega)=\hat{G}(\boldsymbol{x},\omega)&+\frac{\cos(k|\boldsymbol{x}|)}{4\pi^2|\boldsymbol{x}|}\{\mathrm{Si}[(K-k)|\boldsymbol{x}|]+\mathrm{Si}[(K+k)|\boldsymbol{x}|]\\&-\pi\}\\&+\frac{\sin(k|\boldsymbol{x}|)}{4\pi^2|\boldsymbol{x}|}\{\mathrm{Ci}[(K-k)|\boldsymbol{x}|]-\mathrm{Ci}[(K+k)|\boldsymbol{x}|]\}\end{aligned} \tag{23}$$

for $|\boldsymbol{x}|>0$, and

$$\hat{G}_K(\boldsymbol{x},\omega)=\frac{K}{2\pi^2}+\frac{k}{4\pi^2}\left[\ln\left(\frac{K-k}{K+k}\right)-j\pi\right] \tag{24}$$

for $|\boldsymbol{x}|=0$. In Eq. (23), $\mathrm{Si}(x)$ and $\mathrm{Ci}(x)$ are the sine and cosine integrals.[22]

Next, the temporal windowing is performed in the temporal Fourier domain. For the *unfiltered* Green's function, this yields

$$\begin{aligned}G^T(\boldsymbol{x},t)&=\frac{\delta(t-|\boldsymbol{x}|/c_0)}{4\pi|\boldsymbol{x}|}[H(t+T)-H(t-T)]\\&=\frac{\delta(t-|\boldsymbol{x}|/c_0)}{4\pi|\boldsymbol{x}|}H(T-|\boldsymbol{x}|/c_0),\end{aligned} \tag{25}$$

and therefore

$$\hat{G}^T(\boldsymbol{x},\omega)=\frac{\exp(-jk|\boldsymbol{x}|)}{4\pi|\boldsymbol{x}|}H(T-|\boldsymbol{x}|/c_0). \tag{26}$$

This result can be applied directly to the first term of Eq. (23), i.e., the unfiltered Green's function. The time windowing of the other terms in Eq. (23), which represent the spatial filtering operation, is more involving. When the time windowing of these terms is simply omitted, a small time-domain aliasing error is made during the evaluation of the temporal convolution sum. As explained at the end of Sec. III C, the temporal filtering of the term $G^T(\boldsymbol{x},\omega)$ that must still take place is going to yield a comparable time-domain aliasing error, so it is decided to restrict the time windowing to the first term of Eq. (23) and to leave the other terms in this equation unaffected. The remaining temporal filtering and spatial windowing operations are performed in the temporal Fourier domain as well by multiplying $\hat{G}_K^T$ with the appropriate rectangular windows. This results in $\hat{G}_{K,X}^{T,\Omega}(\boldsymbol{x},\omega)$.

C. Spatiotemporal filtering of the primary sources

In all practical cases, the spatial support of the primary source $S^{(0)}(\boldsymbol{x},t)$ is limited and the spatial windowing is unnecessary. Likewise, pulsed sources usually have a limited temporal support that makes the temporal windowing superfluous.

When the source signature has a temporal spectrum that is insignificant beyond Ω, the temporal filtering may be skipped as well without the danger of introducing noticeable aliasing. In many other situations, the Fourier transform of the signature is known in analytical form, so the temporal filtering may conveniently be performed in the temporal Fourier domain. The spatial filtering may be applied in a

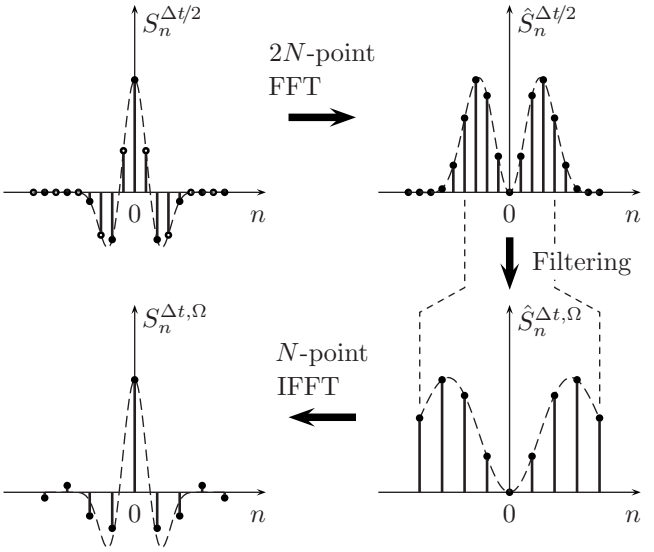


FIG. 2. Numerical procedure for filtering. In this example, the intended number of gridpoints is $N=9$ and the oversampling factor is $a=2$. In the upper left graph, the intermediate samples are indicated by open circles. The dashed curves apply to the unfiltered continuous function.

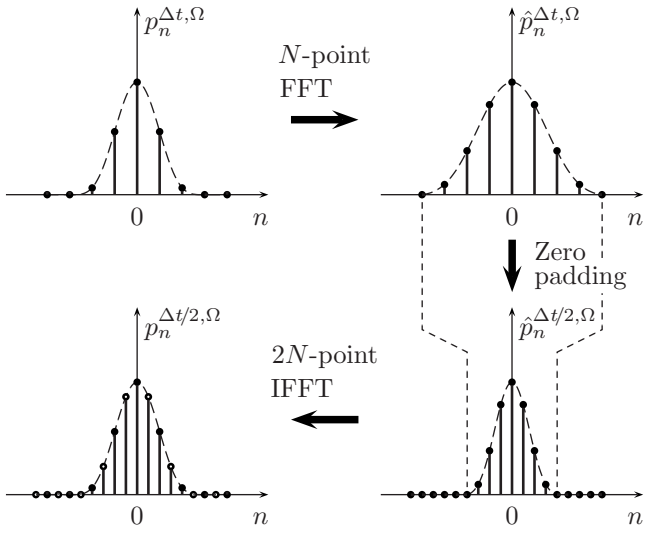


FIG. 3. Numerical procedure for interpolation. In this example, the available number of gridpoints is $N=9$ and the oversampling factor is $a=2$. In the lower left graph, the intermediate samples are indicated by open circles. The dashed curves apply to the continuous function.

similar way whenever the Fourier transform of the source geometry is available in analytical form, e.g., for rectangular or cylindrical piston sources, or for phased array transducers.[23] In such cases, the sampling is directly applied in the Fourier domain. Analytical spatial filtering avoids problems such as staircase effects or the missing out of small sources that are located in between the gridpoints.

When for some coordinate the filtering of the source cannot be employed before the sampling, e.g., when for the relevant coordinate the source can only be characterized through discrete measurements, aliasing must be avoided in a numerical way. For this purpose, the procedure as exemplified in Fig. 2 is proposed. The procedure starts with oversampling the relevant source window at aN points instead of the intended N points. Subjecting the samples to an aN-point FFT yields a discrete version of the Fourier domain function over an interval that is a times as long as intended but, for a properly chosen oversampling factor a, with negligible aliasing at the N points around the origin. Next, all points except those N points around the origin are discarded, thus giving the Fourier domain function over an interval of the intended length. This restriction is the actual filtering step since the remaining discrete interval corresponds with the continuous angular frequency interval $[-\Omega,\Omega]$. Finally, an N-point inverse FFT is performed. This yields N samples in the original domain, at the intended sampling interval and without aliasing. The proposed approach is suitable for the numerical filtering with respect to the spatial coordinates as well.

D. Spatiotemporal filtering of the contrast source

The specific behavior of the contrast source $S^{(j)}(\boldsymbol{x},t)$ depends on the type of contrast present in the medium, as shown at the end of Sec. II A. In any case, $S^{(j)}$ depends on the estimate $p^{(j-1)}$. To discuss the filtering of the contrast source, it is assumed that $p^{(j-1)}$ is bandlimited in all dimensions, with a temporal angular cutoff frequency Ω and a spatial angular cutoff frequency K. Differentiation or convolution of a bandlimited signal yields a result with a bandwidth not exceeding that of the original signal. However, multiplication of a bandlimited signal with another bandlimited signal or with itself causes a result with a bandwidth that is equal to the sum of the bandwidths of both signals involved. Thus, in case of dispersion [Eq. (5)], the contrast source does not require additional filtering. But in case of inhomogeneity [Eq. (4)], the contrast source must be filtered in the spatial dimensions, and in case of nonlinearity [Eq. (6)], the source must be filtered in both the spatial and the temporal dimensions. Detailed analysis shows that for nonlinearity the temporal filtering usually suffices and the spatial filtering may be skipped, but this issue is beyond the scope of this paper.

The abovementioned facts are true for continuous functions as well as their sampled counterparts. Usually, only the sampled version of $p^{(j-1)}$ is available and $S^{(j)}$ must be determined numerically. When this requires the sampled version of a contrast function, e.g., $[c_0^{-2}-c^{-2}(\boldsymbol{x})]$, this must be filtered either analytically (before the sampling) or numerically (after the sampling) to avoid aliasing. When the sampled function $p^{(j-1)}$ is multiplied with a sampled contrast function or with itself, the original sampling intervals $\Delta t=\pi/\Omega$ and $\Delta x=\pi/K$ are no longer sufficient to deal with the increased bandwidth of the result. To avoid aliasing, the relevant sampling intervals must be reduced by some factor a before the multiplication or the squaring takes place. This requires the interpolation of the available samples of $p^{(j-1)}$ without the distortion of the corresponding spectral values. To achieve this, the procedure as depicted in Fig. 3 is suggested. First, an N-point FFT is applied to the available N sample points. The discrete Fourier domain is then extended from N to aN points by zero-padding. Finally, an aN-point inverse FFT is performed. This yields aN samples in the original domain, without distorting the relevant spectral values.

Using the interpolated samples of $p^{(j-1)}$, the relevant

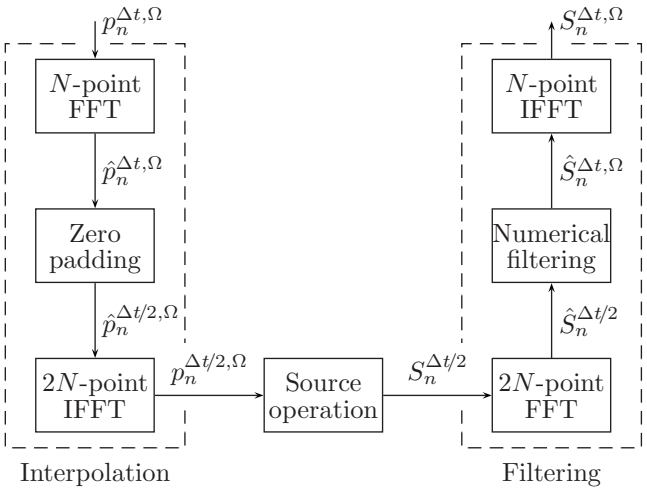


FIG. 4. The entire numerical procedure for the evaluation of the contrast source.

source operation yields aN samples of $S^{(j)}$. This oversampled $S^{(j)}$ is directly amenable for a subsequent numerical filtering according to Fig. 2. The entire numerical procedure for the evaluation of the contrast source is summarized in Fig. 4. At this point, the evaluation of the filtered contrast source may seem computationally involving. However, the contrast source very seldom requires filtering in all four dimensions. When filtering is only required for a subset of the dimensions, the number of operations required for the filtering is at least one order less than for the four-dimensional convolution, and it always pays to reduce the number of grid points by using a proper filtering.

E. Discretization and evaluation of the convolution sum

The filtered and windowed Green's function from Sec. IV B allows for a straightforward sampling. In Secs. IV C and IV D, it has been explained how sampled versions of the filtered and windowed primary source and contrast source are obtained. Apart from the small error introduced by combining filtering and windowing, all aliasing errors have been removed and therefore the sampling may be taken as coarse as $D_\Phi=2$ in all dimensions. The resulting convolution sum can be evaluated efficiently by using a four-dimensional FFT method. To avoid wraparound, all source intervals must be doubled in length by zero-padding, as explained in Sec. III B. As a result, the number of operations required for the convolution is of order $16N^4 \log(16N^4)$.

V. PARALLEL IMPLEMENTATION

The filtered convolution method has been implemented on a single processor system as well as on a multiprocessor parallel system with distributed memory. In both cases, the code has been written in FORTRAN. This section focuses on the parallel implementation[24] of the method as it occurs in the context of the Neumann iterative solution. To ease the notation, the filtering and the windowing of the acoustic quantities are understood without specific indication.

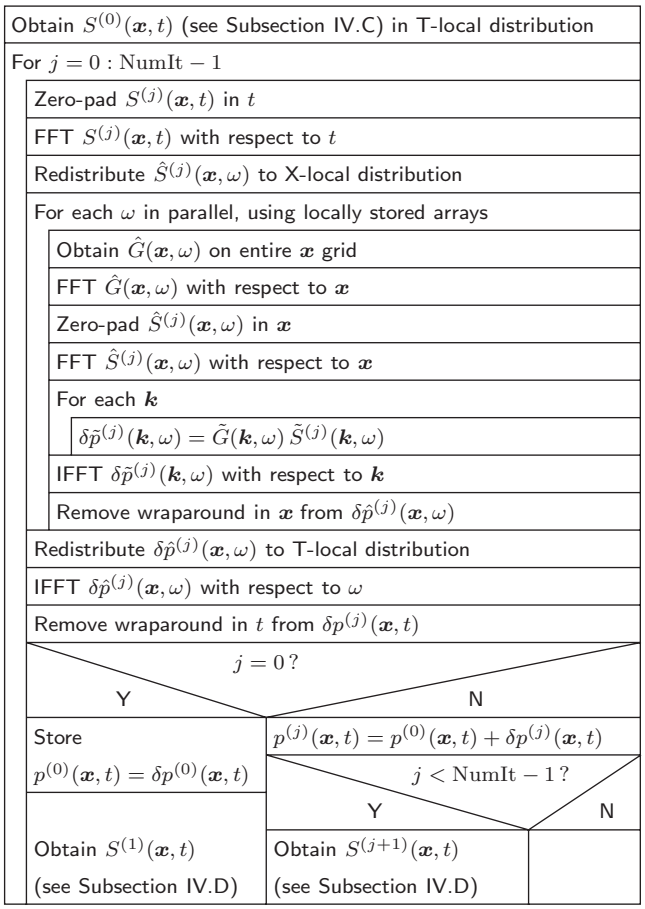


FIG. 5. Structure diagram of the parallel program. All symbols indicating acoustic quantities refer to the filtered and windowed quantities. Note that $\delta p^{(j)}=\mathcal{L}[S^{(j)}]$.

The structure diagram of the parallel program is shown in Fig. 5, in which the evaluation of the convolution plays a central role. To minimize communication, depending on the operation to be applied, either a T-local distribution is used, where each processor is allocated a complete time trace for a limited number of spatial grid points, or an X-local distribution is used, where each processor is allocated the entire spatial domain for a number of time instances or angular frequencies. Inside the outer loop over j, the source function $S^{(j)}(\boldsymbol{x},t)$ is zero-padded in time, transformed in the temporal dimension, and redistributed to X-local. This way, inside the nested loop over ω, the local temporary arrays that hold the spatial data including the zero-padding regions need only be three-dimensional. Another advantage of this approach is that the redistribution steps occur outside the loop over ω. Inside this loop, $\hat{G}(\boldsymbol{x},\omega)$ is evaluated for each spatial point. Either a high-order finite difference scheme or a spectral derivative[25] is used for the evaluation of the temporal derivatives that occur in the primary source function $S^{(0)}(\boldsymbol{x},t)$ and in the contrast source function $S^{(j)}(\boldsymbol{x},t)$ (see Sec. II A). Especially when the numerical interpolation of Fig. 3 is invoked, the route via the transform domain yields the spectral derivative at negligible extra cost. When this option is chosen, care is taken to avoid Gibb's phenomenon[20] that might arise from a jump between the data at the beginning and at the end of the interval that will be transformed by the FFT.

As the filtered convolution method relies heavily on the use of multi-dimensional FFT's, a fast and flexible implementation of the FFT is essential. This is achieved by using the FFTW library.[26]

VI. NUMERICAL RESULTS

The results in this section have been obtained by running the parallel program on an SGI Altix 3700 clustered multiprocessor system of Itanium 2 processors running at 1.3 GHz and with 2 Gbyte memory per processor.

A. The primary source

In the examples that follow, the primary source is flat and located in the plane $z=0$. The source function either follows from a volume density of injection rate $q=Q(x,y,t)\delta(z)$, giving $S=\rho_0\partial_t Q(x,y,t)\delta(z)$, or from a volume density of force $f_z=F_z(x,y,t)\delta(z)$, giving $S=-\partial_z[F_z(x,y,t)\delta(z)]$. The action of these respective sources will be represented by a jump Δv_z in the z component of the particle velocity or a jump Δp in the acoustic pressure. These jumps are only nonzero at the location of the source aperture. For the emitted wave field, symmetry properties make that for $z>0$ these source/saltus descriptions relate to the boundary value descriptions of a source in a perfectly rigid baffle and in a perfectly compliant baffle, respectively, by taking $Q(x,y,t)=2v_{z0}(x,y,t)$ or $F_z(x,y,t)=2p_0(x,y,t)$, where v_{z0} and p_0 are the normal velocity and the pressure prescribed on the boundary. The source functions $Q(x,y,t)$ and $F_z(x,y,t)$ are separated into a time signature and a geometry factor $0\leq A(x,y)\leq 1$. For the time signature, a harmonic wave modulated by a Gaussian envelope is taken. These choices give

$$\begin{pmatrix} Q(x,y,t) \\ F_z(x,y,t) \end{pmatrix} = \begin{pmatrix} Q_0 \\ F_0 \end{pmatrix} A(x,y)\exp[-(2t/t_w)^2]\sin(\omega_0 t), \quad (27)$$

where Q_0 and F_0 are the peak source values, t_w is the pulse width, and ω_0 is the angular center frequency. If there are any position-dependent delays of the time signature, as with focused sources, these may be accounted for in the temporal Fourier domain by using a complex geometry factor

$$\hat{A}(x,y,\omega) = A(x,y)\exp[-j\omega t_d(x,y)], \quad (28)$$

where $t_d(x,y)$ is the delay time. With regard to the temporal filtering of S, the pulse width t_w is taken such that the significant part of the spectrum always occurs below the cutoff frequency, and the filtering operation is omitted. Spatial filtering in the x and y dimensions is performed with the numerical method of Fig. 2. In the z dimension, filtering of the delta function with an ideal filter gives

$$\frac{K}{\pi}\delta(z) *_z \operatorname{sinc}(Kz) = \frac{1}{\Delta x}\operatorname{sinc}(\pi z/\Delta x), \quad (29)$$

which for $z=n\Delta x$ is $1/\Delta x$ for $n=0$ and 0 for $n\neq 0$. In case of a force source, the spatial derivative with respect to z is not directly applied to the delta function but to the result of the convolution, according to the property $G*_z(\partial_z f_z)=\partial_z(G*_z f_z)$.

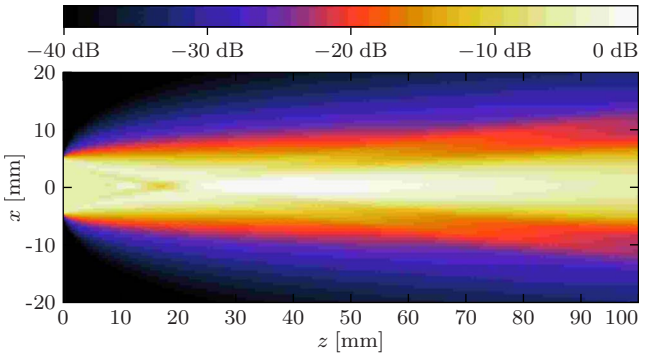


FIG. 6. (Color online) Pressure profile in the plane $y=0$ mm for the cylindrical piston transducer in water. The 0 dB level corresponds to 19.4 kPa. The sampling is $D_F=2.33$.

B. Cylindrical piston transducer in a homogeneous medium

The first example involves a cylindrical piston transducer with a radius of 5 mm, exciting a pulse with a center frequency $f_0=2$ MHz, a pulse width $t_w=3/f_0$, and a maximum surface pressure $p_0=10$ kPa. The action of the transducer is represented by a pressure jump condition. The transducer is spatially filtered by oversampling with a factor $a=8$ followed by numerical filtering, and the derivative in the source function is evaluated with a 30th-order finite difference scheme. The acoustic medium is water with $\rho_0=998$ kg m^{-3} and $c_0=1480$ m s^{-1}. For the applied pressures and frequencies, this medium may be considered linear and lossless. The homogeneous field solution $p^{(0)}(\boldsymbol{x},t)$ has been determined in a computational domain with an approximate spatial size of $40\times 2\times 100$ mm^3 and a co-moving time frame of duration $20/f_0$. The discretization of this domain is given by $D_F=1/F\Delta t$, where the maximum frequency of interest $F=\Phi/2\pi$ is taken 3 MHz. At this frequency, the spectrum of the excitation pulse is 50 dB below the level of the center frequency. Figure 6 shows the computed pressure profile, i.e., the maximum temporal acoustic pressure, in the plane $y=0$ mm for $D_F=2.33$.

To assess the accuracy of the results, in the plane $y=0$ mm the relative root mean square (RRMS) error with respect to a reference solution p_{ref} is considered. This reference follows from a Lobatto integration of an exact expression for the field of the pressure jump source, as derived with the coordinate transformation method described by Harris.[27] The RRMS error is given by

$$\text{Err} = \left(\int_{\mathcal{V}} (p^{(0)} - p_{\text{ref}})^2 dxdzdt \Big/ \int_{\mathcal{V}} (p^{(0)})^2 dxdzdt \right)^{1/2}, \quad (30)$$

where $\mathcal{V}$ is the total xzt-domain. The integrals are evaluated using the left Riemann sum. Table I presents the RRMS error and some other information for different values of D_F. These numerical experiments show that the results for $D_F=2.33$ compare quantitatively very well with the reference solution. The arrival times of the body and edge waves are reproduced correctly. This is also the case for $D_F=2.00$, although for this discretization the error is 4%. The primary cause for this

TABLE I. Data for several computations involving the cylindrical piston transducer in water. The computations are performed for different values of D_F.

D_F	Err (%)	4D grid size ($\times 10^6$ points)	Wall clock time (s)	Number of processors
2.00	4.0	25.72	143	4
2.33	1.0	39.26	230	4
2.67	0.8	60.24	393	4
3.33	0.5	155.7	873	4
4.00	0.5	345.6	686	8
5.33	0.4	1150	1203	48

error is the high-order finite difference scheme that is used for the spatial derivative in the source function. For higher values of D_F, the error decreases quickly to 0.4% at $D_F=5.33$. The main contribution to the latter error comes from the spatial filtering, which is most prominent at the first plane beyond the source. If the results for this plane are left out, the errors for D_F being equal to 3.33, 4.00, and 5.33 decrease to 0.2%, 0.1%, and 0.1%, respectively.

The example proves that domains up to approximately 500×10^6 grid points can be handled successfully and with a limited number of processors. This number of grid points corresponds to a four-dimensional computational domain of 75 wavelengths/periods in each dimension. The latter observation clearly indicates the applicability of the filtered convolution method to very large-scale problems.

As can be seen from the last line of Table I, the problem size and computation time grow massively for increasing values of D_F. This illustrates once more the necessity of using a coarse discretization.

C. Phased array transducer in a homogeneous medium

The second example concerns a 64 element phased array transducer with elements measuring 0.25 mm (width in the x direction) by 12 mm (height in the y direction), and a pitch of 0.3 mm. The elements excite the same surface pressure as the cylindrical transducer mentioned above, and the array is focused at $z=60$ mm, including elevation focusing. The results of the current method are compared with the results from the FIELDII program.[28] Since this program solves the Rayleigh integral,[27] the action of the transducer is now represented by a velocity jump. The phased array source is spatially filtered in the x dimension by analytically filtering one element and subsequently stacking the results, and in the y dimension by numerically filtering a single element by the method of Fig. 2. The temporal derivative in the source function is performed analytically. Again, the medium is water. The homogeneous field solution $p^{(0)}(\boldsymbol{x},t)$ has been determined in a computational domain with an approximate spatial size of $80\times 1\times 100$ mm^3 and a co-moving time frame of duration $40/f_0$. Figure 7 shows the computed pressure profile in the plane $y=0$ mm for $D_F=2.00$ with $F=3$ MHz.

Table II presents the RRMS error in the plane $y=0$ mm, together with other data relating to several discretizations. The listed error has been determined relative to the solution of FIELDII, which was run with a sampling frequency of 400 MHz and 20×40 mathematical elements per physical element. Table II shows that already for $D_F=2.00$ the results from both programs compare very well. This clearly proves that the filtered convolution method is capable of accurately computing the space-time domain acoustic field of complex plane sources, even when sampled at the limit $D_F=2$.

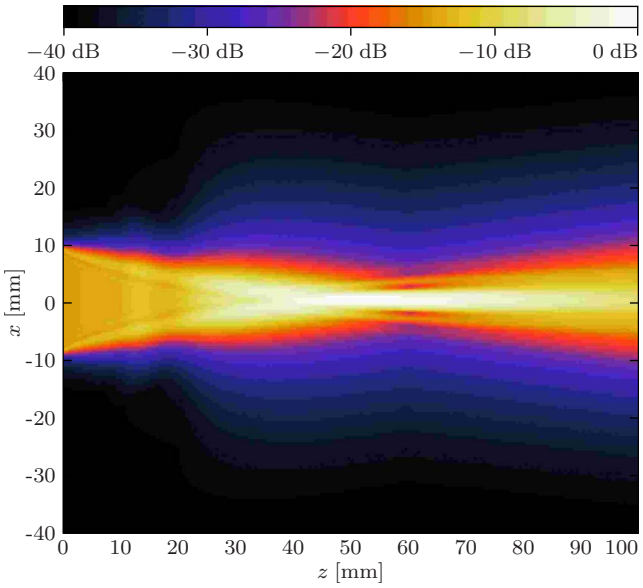


FIG. 7. (Color online) Pressure profile in the plane $y=0$ mm for the phased array transducer in water. The 0 dB level corresponds to 47 kPa. The sampling is $D_F=2.00$.

D. Phased array transducer in an inhomogeneous medium

The third example concerns a phased array transducer with the same geometry and surface pressure as in Sec. VI C. This transducer, however, excites an acoustic field in a medium with an inhomogeneous wave speed and a constant mass density. The distribution of the wave speed in the medium is displayed in Fig. 8(a). The homogeneous embedding is water. This surrounds a homogeneous phase screen ($c=0.95c_0$) with an undulation in the x direction, a water-filled cylinder ($c=0.95c_0$) with a homogeneous wall, and a prolate ellipsoid with an aspect ratio of 2 and a continuously changing wave speed ($\max[c]=1.10c_0$). The inhomogeneous wave speed leads to the contrast source of Eq. (4). The contrast function $[c_0^{-2}-c^{-2}(\boldsymbol{x})]$ is filtered by using the method of Fig. 2 with an oversampling factor $a=2$. Successive approximations $p^{(j)}$ up to $j=9$ have been determined in a computational domain with a spatial size of $42\times 16\times 60$ mm^3 and a

TABLE II. Data for several computations involving the phased array transducer in water. The computations are performed for different values of D_F.

D_F	Err (%)	4D grid size ($\times 10^6$ points)	Wall clock time (s)	Number of processors
2.00	1.6	67.0	352	4
2.67	1.0	161.6	510	8
3.33	0.7	318.1	1380	8

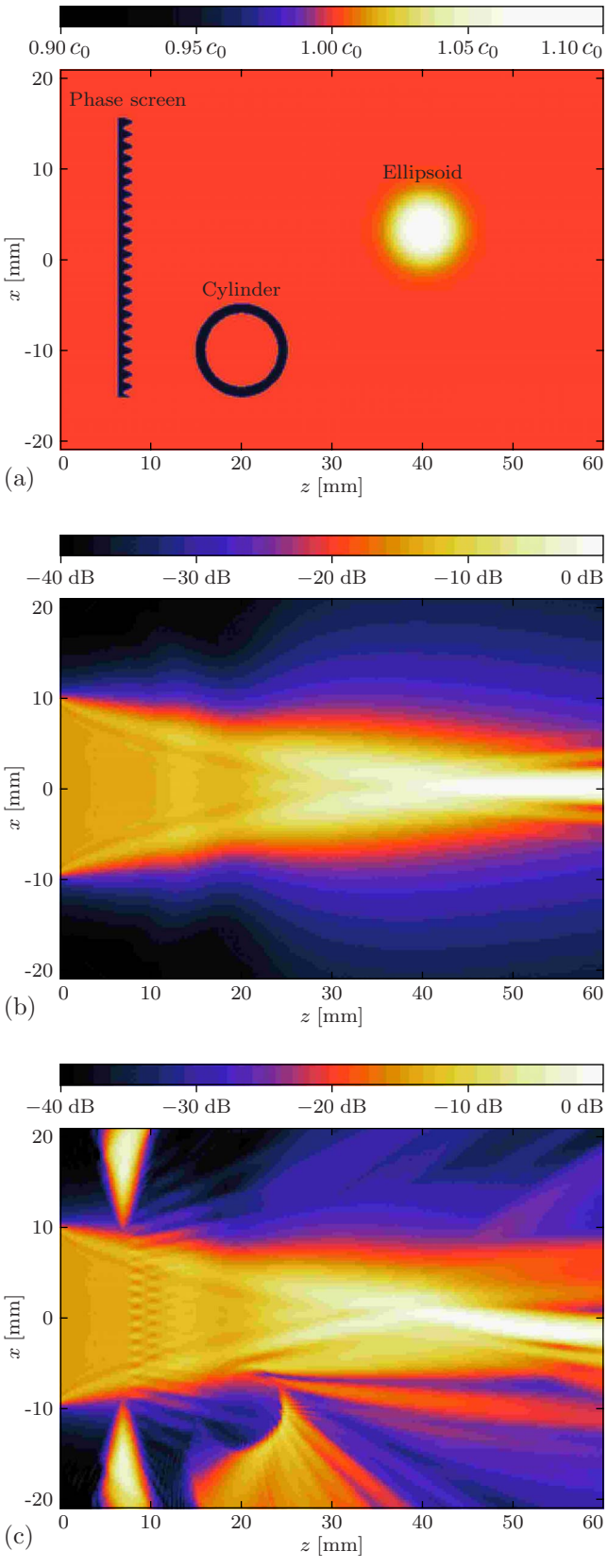


FIG. 8. (Color online) The relative wave speed (a), the pressure profile of $p^{(0)}$ (b), and the pressure profile of $p^{(9)}$ (c) in the inhomogeneous medium, for the plane y=0 mm. In panels (b) and (c), the 0 dB level corresponds to 47 kPa and the sampling is D_F=2.66.

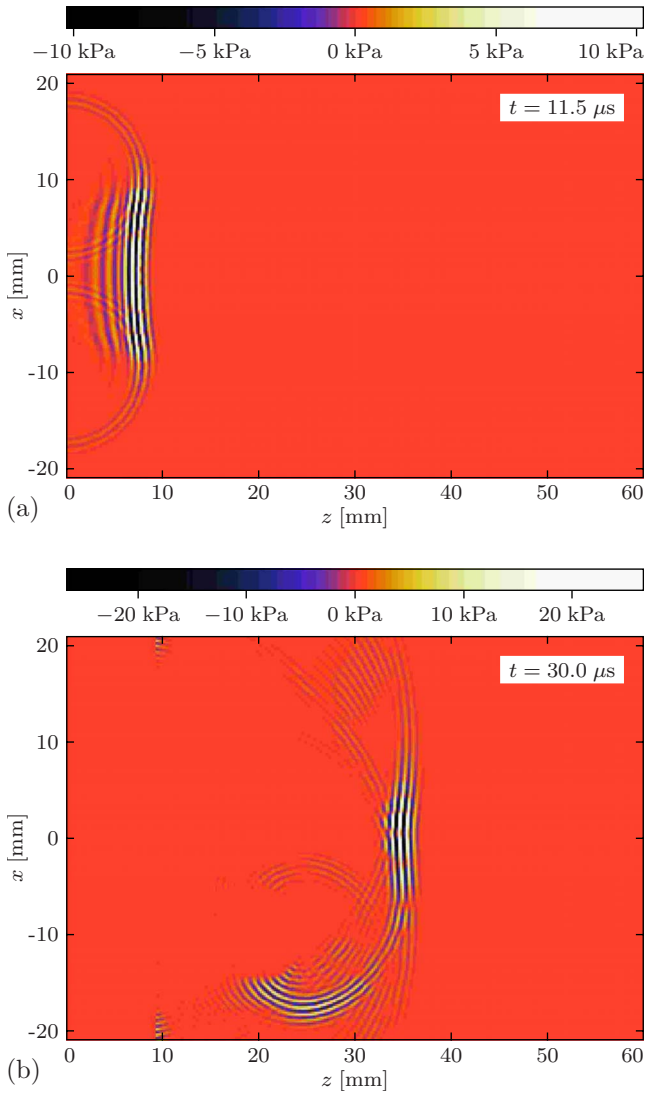


FIG. 9. (Color online) Snapshots of $p^{(9)}$ in the inhomogeneous medium, for the plane y=0 mm. The sampling is D_F=2.66.

co-moving time frame of duration $48/f_0$. Figures 8(b) and 8(c) show the profiles of $p^{(0)}$, i.e., the incident field, and $p^{(9)}$, i.e., an approximation of the total field, in the plane y=0 mm.

The results in Fig. 8(c) clearly reveal the effect of the objects on the beam shape. The phase screen affects the beam shape by generating an interference pattern just beyond the screen and by forming two side-lobes. The wall of the cylinder acts as a curved waveguide, which captures a part of the beam and radiates it off in the negative x direction. The main effect of the ellipsoid, which acts as a diverging acoustic lens, is the diffraction of the beam in the negative x direction. In Fig. 9, which shows snapshots of $p^{(9)}$, the reflections and interferences due to the inhomogeneities are also clearly visible.

In the current case, the successive results are almost identical from j=5 to j=9. However, due to the lack of benchmarking results, it has neither been investigated whether in the given situation the Neumann iterative solution is convergent, nor whether a possible convergence is toward the correct result.[29] In any case, the filtered convolution

method may be combined with more sophisticated iterative schemes[3] with improved convergence properties, such as the successive over-relaxation method or the conjugate gradient method.

The results above have been obtained for $D_F=2.66$ with $F=3$ MHz. For the computation of the successive approximations, the entire spatial domain of $85\times32\times121$ wavelengths and a co-moving time interval of 72 periods, all relating to F, acted as the support of the contrast source. The spatiotemporal grid consisted of 1.16×10^9 points, and it took 4.3 wall clock hours with 48 processors to obtain the iterations up to $j=9$. This proves that very large-scale, complex acoustic wave fields can efficiently be computed with the filtered convolution method.

VII. CONCLUSIONS

A method has been developed for the numerical evaluation of convolutions over very large, four-dimensional domains. The method is quite general and may, for instance, be combined with various iterative integral equation methods for the solution of numerous contrast source problems. In this role, the presented method enables the computation of very large spatiotemporal wave fields by minimizing the storage requirement and strongly reducing the computational effort.

As commonly encountered, the numerical convolutions are performed with the aid of FFT's. The distinguishing feature of the method is that before sampling, the relevant functions are systematically filtered, windowed, and zero-padded with respect to all relevant coordinates. Both analytical and numerical filtering procedures have been presented. Due to these preparatory steps, it is possible to perform numerical convolutions with a sampling that approaches the limit $D_\Phi=2$, i.e., two points per wavelength and per period of the highest angular frequency of interest Φ.

The method has been combined with the Neumann iterative solution of the acoustic contrast source problem for an inhomogeneous medium. Within this context, the method has been implemented on a parallel computer, and numerical results have been obtained for several configurations. Comparison with two separate reference solutions for a homogeneous medium shows that the RRMS error of our results is a few percent for $D_\Phi=2$, and decreases rapidly for finer sampling. The major part of this error is due to the filtering operation, and is thus made consciously. Computation of the transient acoustic wave field in an inhomogeneous medium proves that with the presented method, computational domains in the order of 100 wavelengths in three spatial directions by 100 periods in time can conveniently be dealt with.

ACKNOWLEDGMENTS

This research was supported by the Dutch Technology Foundation (STW) and the Dutch National Computing Facilities Foundation (NCF).

[1]P. van den Berg, "Iterative computational techniques in scattering based upon the integrated square error criterion," IEEE Trans. Antennas Propag. **32**, 1063–1071 (1984).
[2]P. van den Berg, in *PIER 5: Application of Conjugate Gradient Method to Electromagnetics and Signal Analysis*, edited by T. Sarkar (Elsevier, New York, 1991), pp. 27–65.
[3]R. Kleinman and P. van den Berg, in *PIER 5: Application of Conjugate Gradient Method to Electromagnetics and Signal Analysis*, edited by T. Sarkar (Elsevier, New York, 1991), pp. 67–102.
[4]R. Kleinman and P. van den Berg, "Iterative methods for solving integral equations," Radio Sci. **26**, 175–181 (1991).
[5]C. Shannon, "Communication in the presence of noise," Proc. IEEE **86**, 447–457 (1998).
[6]W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1996).
[7]A. Taflove, *Computational Electrodynamics: The Finite Difference Time-Domain Method* (Artech House, Boston, 1995).
[8]J. Jin, *The Finite Element Method in Electromagnetics*, 2nd ed. (Wiley, New York, 2002).
[9]N. Bojarski, "The k-space formulation of the scattering problem in the time domain," J. Acoust. Soc. Am. **72**, 570–584 (1982).
[10]Q. Liu, "Generalization of the k-space formulation to elastodynamic scattering problems," J. Acoust. Soc. Am. **97**, 1373–1379 (1995).
[11]T. Mast, L. Souriau, D.-L. Liu, M. Tabei, A. Nachmann, and R. Waag, "A k-space method for large-scale models of wave propagation in tissue," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 341–354 (2001).
[12]M. Tabei, T. Mast, and R. Waag, "A k-space method for coupled first-order acoustic propagation equations," J. Acoust. Soc. Am. **111**, 53–63 (2002).
[13]P. Morse and K. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[14]A. de Hoop, *Handbook of Radiation and Scattering of Waves* (Academic, San Diego, 1995).
[15]L. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (IEEE, New York, 1994).
[16]G. Barton, *Elements of Green's Functions and Propagation: Potentials, Diffusion and Waves* (Oxford University Press, Oxford, 1989).
[17]J. DeSanto, *Scalar Wave Theory: Green's Functions and Applications* (Springer-Verlag, Berlin, 1992).
[18]J. Fokkema and P. van den Berg, *Seismic Applications of Acoustic Reciprocity* (Elsevier, Amsterdam, 1993).
[19]E. Kamen, *Introduction to Signals and Systems*, 2nd ed. (McMillan, New York, 1987).
[20]A. Poularikas and S. Seely, *Signals and Systems*, 2nd ed. (PWS-KENT, Boston, 1991).
[21]R. van der Veeken and H. Blok, "A configurational filtering and sampling method in the numerical solution of three-dimensional electromagnetic source problems," J. Electromagn. Waves Appl. **4**, 919–943 (1990).
[22]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, 9th ed. (Dover, New York, 1972).
[23]R. Zemp, J. Tavakkoli, and R. Cobbold, "Modeling of nonlinear ultrasound propagation in tissue from array transducers," J. Acoust. Soc. Am. **113**, 139–152 (2003).
[24]A. Grama, A. Gupta, G. Karypis, and V. Kumar, *Introduction to Parallel Computing*, 2nd ed. (Pearson Education Limited, Harlow, England, 2003).
[25]B. Fornberg, *A Practical Guide to Pseudospectral Methods* (Cambridge University Press, Cambridge, 1998).
[26]M. Frigo and S. Johnson, "The design and implementation of FFTW3," Proc. IEEE **93**, 216–231 (2005).
[27]G. Harris, "Review of transient field theory for a baffled planar transducer," J. Acoust. Soc. Am. **70**, 10–20 (1981).
[28]J. Jensen and N. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267 (1992).
[29]The behavior of the Neumann iterative solution in combination with the filtered convolution method has been further investigated for a configuration consisting of an acoustically penetrable, homogeneous sphere in a homogeneous background medium (see Ref. 11). For this case analytical results exist. Successive approximations $\delta p^{(j)}=p^{(j)}-p^{(0)}$ of the scattered field have been determined for several radii of the sphere, several values of the wave speed inside the sphere, and several values of the sampling parameter D_F. For those situations in which the successive approximations clearly tend to a limiting result, the RRMS error of this limiting result with respect to the exact solution has been evaluated. From these comparisons, it may be concluded that the limiting results are indeed an approximation of the exact solution, and become better for higher values of D_F. The latter is explained by the fact that the observed error is almost entirely caused by the spatial filtering of the contrast function of the relatively small sphere.

A micro-machined source transducer for a parametric array in air[a)]

Haksue Lee, Daesil Kang, and Wonkyu Moon
Department of Mechanical Engineering, Pohang University of Science and Technology (POSTECH), San 31, Hyojadong Namgu, Pohang, Kyungbuk 790-784, South Korea

(Received 15 March 2008; revised 15 January 2009; accepted 16 January 2009)

Parametric array applications in air, such as highly directional parametric loudspeaker systems, usually rely on large radiators to generate the high-intensity primary beams required for nonlinear interactions. However, a conventional transducer, as a primary wave projector, requires a great deal of electrical power because its electroacoustic efficiency is very low due to the large characteristic mechanical impedance in air. The feasibility of a micro-machined ultrasonic transducer as an efficient finite-amplitude wave projector was studied. A piezoelectric micro-machined ultrasonic transducer array consisting of lead zirconate titanate uni-morph elements was designed and fabricated for this purpose. Theoretical and experimental evaluations showed that a micro-machined ultrasonic transducer array can be used as an efficient source transducer for a parametric array in air. The beam patterns and propagation curves of the difference frequency wave and the primary wave generated by the micro-machined ultrasonic transducer array were measured. Although the theoretical results were based on ideal parametric array models, the theoretical data explained the experimental results reasonably well. These experiments demonstrated the potential of micro-machined primary wave projector.
 [DOI: 10.1121/1.3081385]

PACS number(s): 43.25.Lj, 43.38.Ar, 43.38.Gy, 43.38.Fx [AJZ] Pages: 1879–1893

I. INTRODUCTION

Parametric arrays (PAs) have been the subject of theoretical[1,2] and experimental[3] studies since the 1960s and are used in active underwater SONAR systems such as sub-bottom profilers[4] or communication transmitters.[5] Their principal advantage is that the highly directional low-frequency sound beam required can be indirectly generated with a relatively small transducer. After an experimental study in air,[6] many researchers and engineers turned their attention to the potential of PA applications in this medium, especially for highly directional parametric loudspeaker systems.[7,8]

Unlike in water, it is difficult to produce the high-intensity acoustic beams required for nonlinear interaction in air with an ultrasonic/acoustic transducer (i.e., source transducer) because the acoustic impedance (415 Pa s/m) is much smaller (1.54 MPa s/m).[9] To achieve the required high-intensity beams with the same transducer in air, the output velocity amplitude should be much higher than in water. Furthermore, the weak nonlinearity and strong attenuation effects in air also reduce the parametric conversion.[6,10] To compensate for the weak radiation characteristics, PA applications in air, especially parametric loudspeakers, have used huge array-type radiators consisting of conventional ultrasonic transducers that require a lot of electrical power.[7] Recently, transducers using active polymers (e.g., piezoelectric or electrostrictive polymers, etc.) have been proposed as high-power radiators in air to improve practical applicability by reducing the volume and simplifying construction.[8,11] One transducer has been applied as a parametric radiator to improve the power efficiency in air, although it still requires a large amount of electrical power.[8] New and useful PA applications, such as small loudspeakers or sensors for mobile systems,[12–14] may be possible if high-intensity source transducers with thinner (or smaller) dimensions and higher efficiency can be developed. Here, the potential for piezoelectric micro-machined ultrasonic transducer (pMUT) arrays to be used as thin, highly efficient, high-intensity primary wave projectors for PAs is examined.

pMUTs and capacitive MUTs have been the objects of numerous studies, mainly for short-range imaging applications such as phased array transducers operating between several and several hundred megahertz in water or air.[15–18] One important feature of thin-film transducers such as MUTs is that their mechanoacoustic efficiency[19] can be improved significantly if the internal mechanical impedance is reduced with thin-film membranes or plates.[20,21] With careful design, this can potentially allow for improvements in the overall electroacoustic efficiency[19] of the transducers because of the improved mechanoacoustic efficiency and low motional loading in air.[16] This feature is important in an air-based acoustic radiator design because the electroacoustic efficiency of bulk piezoelectric lead zirconate titanate (PZT) ceramics is extremely low in air.[22] The main reason for this is due to the mechanoacoustic efficiency and not the electromechanical efficiency.

[a)]Some of these results were presented at the following international conferences: The 151st meeting of the Acoustical Society of America, Providence, Rhode Island (June 5–9, 2006), The SICE-ICASE International Joint Conference 2006, Busan, Korea (October 18–21, 2006), and Transducers and Euroscensors '07, Lyon, France (June 10–14, 2007).

Even though generating high-intensity acoustic wave is generally not too important for micro-machined phased arrays in water, it is possible with micro-machined transducers and should be considered for PAs. There have been reports of pMUT arrays (operating up to hundreds of kilohertz) for generating high-intensity waves in air,[23] but more work on the transducer and array design is still required, especially for PAs in air. The velocity of a micro-machined thin-film radiator should be large enough to generate high-intensity waves. This is expected to be achieved with a proper transduction mechanism design by considering deformation and coupling effects in air. In addition, transducer arrays consisting of small micro-machined transducer units should be designed to generate the required high-intensity sound beams.

It is also important in micro-machined source transducer design to select appropriate operating conditions. These conditions should be based on the physical phenomena involved in the specific application, as well as electroacoustic efficiency and power issues. Although conceptual patents covering micro-machined source transducers and applications in air have recently been filed,[12,13] experimental results have rarely been reported. Although we reported preliminary data[24] and discussed some applications,[24–26] it took us a long time to analyze the experimental results and find the causes of some practical issues[26] encountered. As a fundamental demonstration of capability, a two-channel-type source transducer was designed to generate a difference frequency wave (DFW) with narrow beamwidth to show the feasibility and potential of a micro-machined primary wave projector, and it was evaluated experimentally. The experimental data were compared to theoretical data.

II. SOURCE TRANSDUCER DESIGN

A specific design goal provides coherence to the overall design and experimental procedure. An ultrasonic ranging transducer with a narrow beamwidth in air was selected as a design goal to provide some useful application for PAs in air. The half-power beamwidth (HPBW) of general ultrasonic ranging sensors is too large (18°–50°) (Refs. 27 and 28) to detect the exact location of a small obstacle, and the wide beamwidth often creates low spatial resolution problems in ultrasonic ranging applications such as small mobile robots. Therefore, the source transducer was designed to generate indirectly an ultrasonic pulse with a narrow beamwidth of approximately 5°, chosen to be about the same size as that of conventional ultrasonic ranging sensors.

The source transducer has two major functional requirements. One is to create the proper interaction region to ensure a narrow beamwidth with high sound pressure for the DFW. The other is to generate primary waves of sufficiently high intensity for nonlinear interactions. Therefore, the design procedure consisted in determining the optimal design parameters based on PA theory and constructing the individual pMUTs and transducer array based on those parameters while taking the electroacoustic efficiency into account.

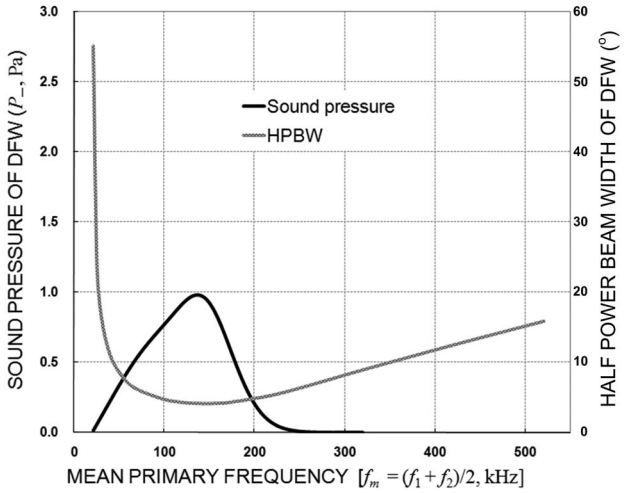
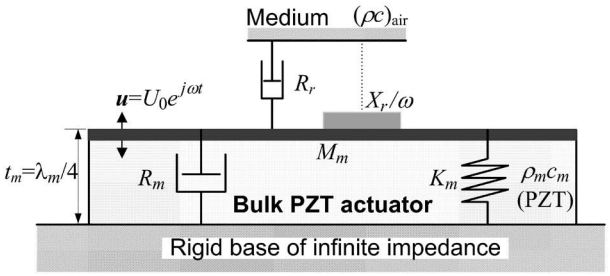


FIG. 1. HPBW and sound pressure of the DFW at 5 m according to the mean frequency of primary waves with selected design parameters (Ref. 30): $2a=32$ mm and $f_d=40$ kHz, assuming an ideal Gaussian source distribution ($P_1=P_2=100$ Pa) with no saturation effects.

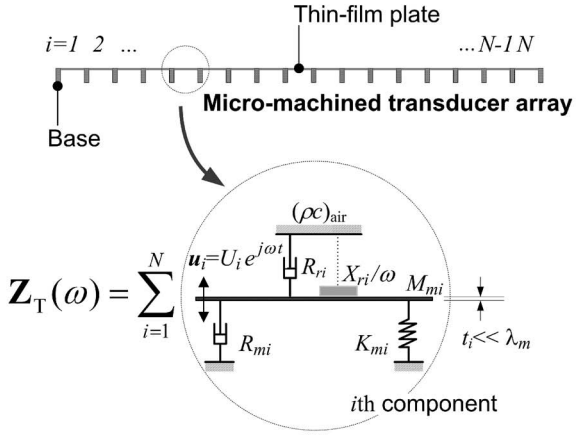
A. Design parameters based on operating conditions

The spatial distribution of the interaction region related to the HPBW and sound pressure level (SPL) of the DFW can be determined by the beam patterns of the primary waves considering diffraction, absorption, and saturation effects.[29,30] Even though these radiation characteristics themselves are complex, previous research[30] has shown that the beamwidth and sound pressure of the DFW can be easily estimated for a given source pressure level of primary waves using the main design parameters: the primary frequencies (f_1 and f_2), the difference frequency (f_d), and the radiation area of the transducer (πa^2, where a is the effective radius of a source transducer).

For this research, a diameter $2a=32$ mm and a difference frequency $f_d=40$ kHz were selected based on the specifications of a conventional ultrasonic ranging sensor. This meant that the primary frequency was the main parameter that determined the beam characteristics of the DFW. Figure 1 shows the HPBW and sound pressure (on the axis of transducer) of the DFW as functions of the mean primary frequency, $f_m=(f_1+f_2)/2$, based on a PA theoretical model.[30] In the calculation,[30] the source transducer was assumed to be a circular radiator with an axisymmetric Gaussian pressure distribution. It should be noted that the model assumed that the f_d/f_m is small, and this means that the graph may not be exactly correct as f_d/f_m becomes larger. The maximum sound pressure and minimum beamwidth could be achieved at an f_m of approximately 140 kHz. Given the selected design values, the expected value of the HPBW was nearly 4°. For this study, the value $f_m=120$ kHz (i.e., $f_1=100$ kHz and $f_2=140$ kHz) was selected instead of the optimal frequency of $f_m=140$ kHz due to the high cutoff frequency (140 kHz) of the microphone used in the experiments. This resulted in a slight increase in the beamwidth and decrease in the sound pressure. The nonlinear interaction length (near-field of the PA) was approximately 2.53 m ($\approx 1/2\alpha_0$, where α_0 is the absorption coefficient at f_m).[6]



(a)



(b)

FIG. 2. Comparison of the equivalent mechanical impedance model. (a) Bulk PZT actuator with quarter-wave resonance (radiation impedance $\mathbf{Z}_r=R_r+jX_r$, internal mechanical impedance $\mathbf{Z}_m=R_m+j(\omega M_m-K_m/\omega)$, and output velocity $\boldsymbol{u}$). (b) Micro-machined transducer array consisting of N identical transducer components (for ith component, radiation impedance $\mathbf{Z}_{ri}=R_{ri}+jX_{ri}$, internal mechanical impedance $\mathbf{Z}_{mi}=R_{mi}+j(\omega M_{mi}-K_{mi}/\omega)$, total mechanical impedance $\mathbf{Z}_T$, and output velocity u_i).

B. Approach to a micro-machined source transducer

Before designing the micro-machined source transducer with the obtained design parameters, the conceptual method for thin-plate flexural transducers to enhance mechanoacoustic efficiency needs to be discussed clearly. Consider the lumped parameter models for a conventional bulk actuator[31] and the micro-machined transducer array, as shown in Fig. 2.[19] For convenience, a matching layer[32] in the bulk actuator is omitted, and the model of the micro-machined clamped circular plate[19] is considered. Before considering the efficiency of the micro-machined transducer array, the mechanoacoustic efficiency of a single driver in the array may be calculated because the efficiency of the array can be estimated by combining the efficiencies of the unit drivers.

For the ith unit driver in Fig. 2(b), the output vibration velocity is $\boldsymbol{u}_i=\mathbf{f}_i/(\mathbf{Z}_{ri}+\mathbf{Z}_{mi})=U_ie^{j\omega t}$ when the harmonic driving force $\mathbf{f}_i=F_ie^{j\omega t}$ excites the driver. Because two energy-consuming components (the internal mechanical damping R_{mi} and acoustical loading R_{ri}) are connected parallel to the mass (i.e., they have the same velocity), the mechanoacoustic efficiency (η_{MA}) of the ith component can be expressed as[19]

$$\eta_{MA,i}=\frac{\text{Radiated power}}{\text{Total dissipated power}}=\Pi_{r,i}/(\Pi_{m,i}+\Pi_{r,i})$$
$$=R_{ri}/(R_{mi}+R_{ri}), \quad (1)$$

where $\Pi_{r,i}$ is the radiated power such that $\Pi_{r,i}=(1/T)\int_0^T\mathrm{Re}[\mathbf{Z}_{ri}\boldsymbol{u}_i]\mathrm{Re}[\boldsymbol{u}_i]dt=\frac{1}{2}R_{ri}U_i^2$ and $\Pi_{m,i}$ is the internal energy loss where $\Pi_{m,i}=(1/T)\int_0^T\mathrm{Re}[\mathbf{Z}_{mi}\boldsymbol{u}_i]\mathrm{Re}[\boldsymbol{u}_i]dt=\frac{1}{2}R_{mi}U_i^2$.

Let a_i and t_i be the radius and thickness of the ith driver, respectively. Because the radiation resistance is at most $1.2(\rho c)_{air}\pi a_i^2$, the mechanical impedance of the driver determines the mechanoacoustic efficiency of the driver, $\eta_{MA,i}$. Because the driver is designed to operate at its first resonance frequency, the lumped parameters for the driver (plate only) can be obtained easily, such that the resonance frequency can be matched as follows:[19,33]

$$K_{mi}=\frac{16\pi}{a_i^2}\cdot\frac{Et_i^3}{(1-\nu^2)}, \quad (2)$$

$$M_{mi}=\frac{192}{\Lambda_0{}^2}\cdot\rho\cdot(\pi a_i^2t_i), \quad (3)$$

where E, υ, ρ, and Λ_0 are the elastic modulus, Poisson's ratio, mass density of the plate, and frequency constant, respectively. The first mode resonance frequency of the circular plate is given as[19]

$$\omega_0=\Lambda_0\frac{t_i}{a_i^2}\left[\frac{E}{12\rho(1-\nu^2)}\right]^{1/2}=\left[\frac{K_{mi}}{M_{mi}}\right]^{1/2}, \quad (4)$$

where $\Lambda_0=10.22$. It is useful to introduce the concept of the characteristic mechanical impedance[34] of a thin-plate driver as follows:

$$Z_{ci}=(M_{mi}K_{mi})^{1/2}=\frac{32\pi}{\Lambda_0}\left[\frac{\rho E}{(1-\nu^2)}\right]^{1/2}t_i^2. \quad (5)$$

In addition, from the definition of the mechanical quality factor, that of the unit driver may be expressed as follows:[31]

$$Q_{mi}=2\pi\times\frac{\text{Energy stored in the vibrator}}{\text{Energy dissipated internally per cycle}}$$
$$=\frac{Z_{ci}}{R_{mi}}=\frac{(M_{mi}K_{mi})^{1/2}}{R_{mi}}. \quad (6)$$

In general, the mechanical Q-factor of an unloaded simple vibration system such as a PZT disk equals that of its material, PZT, and the Q_m of PZT is usually given as a material property.[34]

Now consider the internal mechanical resistance, R_{mi}, in Eq. (6). Because the unit driver is composed of two materials, silicon and PZT, and because there is no joint throughout its structure, its quality factor, Q_{mi}, may be close to those of its component materials, which have values more than 1000 for silicon and about 500 for PZT. In other words, it may be taken as a given value for the unit driver considered here. Hence, the mechanical resistance of the unit driver can be estimated using Eq. (6) as follows:

TABLE I. Material properties.

	In design and simulation		PZT (52/48) film applied in the fabrication[c]			
Quantity	Silicon[a]	PZT 5H[b]	Value	Thickness (μm)	Measurement method	Unit
ε_r	—	3400	1400	1	C-V measurement	—
tan δ	—	0.02	0.02	1		—
d_{33} (film: $d_{33,f}$)	—	593	250	1	Atomic force microscopy	pm/V
d_{31} (film: $d_{31,f}$)	—	−274	−86.9	0.4	Four point bending	pm/V
E	130.2	Anisotropic	N/A			GPa
ρ	2300	7500	N/A			kg/m^3
ν	0.278	Anisotropic	N/A			—

[a]Assume homogeneous, 100 wafer and out-of-plane motion, Ref. 50.
[b]Using anisotropic matrix, Ref. 34.
[c]Data from manufacturer, Ref. 40.

$$R_{mi} = \frac{1}{Q_{mi}} \cdot (M_{mi}K_{mi})^{1/2} = \frac{1}{Q_{mi}} \frac{32\pi}{\Lambda_0} \left[\frac{\rho E}{1-\upsilon^2} \right]^{1/2} \cdot t_i^2. \quad (7)$$

Equation (7) clearly shows that the mechanical resistance of the unit driver decreases with its thickness t_i. In other words, for any given operating frequency, a smaller value of t_i guarantees a smaller value of R_{mi}, whereas the choice of t_i automatically determines the radius of the unit driver a_i. This suggests that the mechanoacoustic efficiency, $\eta_{MA,i}$, increases as the thickness of the unit driver, t_i, decreases according to Eq. (1). Most acoustic drivers in air have R_m and Z_c considerably larger than the radiation resistance, R_r, and R_m is determined by the characteristic mechanical impedance Z_c of the driver because Q_m is a constant that is usually much larger than 1 for a vibrator structure made of the given materials. This implies that Z_c needs to be reduced. However, it is difficult to make a driver of the conventional size and shape in air such that its Z_c is sufficiently small that R_m is close to R_r. However, as Eq. (7) shows, the micro-machined ultrasonic driver may have a sufficiently small Z_{ci} that its R_{mi} is even smaller than the maximum R_{ri} if the plate is sufficiently thin.

Now consider in detail the method used to improve mechanoacoustic efficiency. In the case of the driver considered above, a very high $\eta_{MA,i}$ in air can be achieved with a unit driver of realizable size and materials. Let us choose $\eta_{MA,i}$ to be 80% in air and assume that the driver will be made of silicon (using the properties in Table I) with $Q_{mi}=500$ for the overall structure of the unit driver. The radiation resistance of the driver may be set as $R_{ri} \approx \pi a_i^2 (\rho c)_{\text{air}}$. From Eqs. (1) and (7), it is easily seen that the ratio R_{mi}/R_{ri} determines $\eta_{MA,i}$. Hence, for a given $\eta_{MA,i}=0.8$, the following is easily obtained:

$$\frac{R_{mi}}{R_{ri}} \approx \frac{1}{Q_{mi}} \cdot \frac{32}{\Lambda_0} \left[\frac{\rho E}{1-\upsilon^2} \right]^{1/2} \cdot \frac{1}{(\rho c)_{\text{air}}} \cdot \left(\frac{t_i}{a_i} \right)^2 \approx \frac{1-\eta_{MA,i}}{\eta_{MA,i}} = 0.25. \quad (8)$$

The aspect ratio (t_i/a_i) is determined automatically as $t_i/a_i \approx 0.03$, and it yields $\omega_0 t_i \approx 21.25$ using Eq. (4). Subsequently, the thickness of the unit driver is also determined as $t_i \approx 28$ μm because ω_0 is given by the operating frequency $f_0=120$ kHz $(=f_m)$. This implies that a_i should be $a_i \approx 933$ μm because the aspect ratio (t_i/a_i) is already given. The resulting values for a_i and t_i are typical sizes of a structure that can be fabricated easily using micro-machining technology.

To increase the efficiency of the source transducer, one must apply the previous results to calculate the efficiency of the array of unit drivers. When the transducer array is driven at the resonance frequency of its unit driver, the total mechanoacoustic efficiency $(\eta_{MA,T})$ of the array can be written as

$$\eta_{MA,T} = \sum_{i=1}^{N} \Pi_{r,i} \Big/ \sum_{i=1}^{N} (\Pi_{m,i} + \Pi_{r,i}) = \sum_{i=1}^{N} \frac{1}{2} R_{ri} U_i^2 \Big/ \left(\sum_{i=1}^{N} \frac{1}{2} R_{mi} U_i^2 + \sum_{i=1}^{N} \frac{1}{2} R_{ri} U_i^2 \right). \quad (9)$$

When generating a plane wave, by assuming equal amplitudes with in-phase driving $(\mathbf{f}_i = F_i e^{j\omega t} = F e^{j\omega t}$, $i=1,2,3,\ldots,N)$ of each component and the same acoustical load on each component (ignoring mutual acoustic impedance[35]), one ensures that the output vibration velocity of each component is the same (i.e., $\boldsymbol{u}_i = U_i e^{j\omega t} = U e^{j\omega t}$, $i=1,2,3,\ldots,N$). Thus Eq. (9) can be rewritten as

$$\eta_{MA,T} = \sum_{i=1}^{N} R_{ri} \Big/ \left(\sum_{i=1}^{N} R_{mi} + \sum_{i=1}^{N} R_{ri} \right) = N \cdot R_{ri} / (N \cdot R_{mi} + N \cdot R_{ri}). \quad (10)$$

If t_i is chosen to be as small as possible, as explained above, the total internal mechanical damping $(N \cdot R_{mi})$ of the silicon-based microstructure becomes much lower than that of a conventional thickness mode transducer, which is determined by a given operating resonance frequency. The mechanical resistance of a conventional thickness mode transducer can be expressed as $R_m \approx 1/Q_m \cdot \omega_0 \cdot M_m = 1/Q_m \cdot \omega_0 \cdot \rho_m A t_m$ and t_m is determined as $\lambda_m/4 (=1/4 \cdot c_m/f_0)$ by the given operating resonance frequency. However, the acoustical loading of the transducer array $(N \cdot R_{ri})$ is approximately equal to that of a conventional transducer (R_r) if the whole shape and radiation area are designed to be similar $(N \cdot \pi a_i^2 \approx A)$. This implies that a micro-machined transducer array can achieve a much larger value of $\eta_{MA,T}$ than a conventional thickness mode

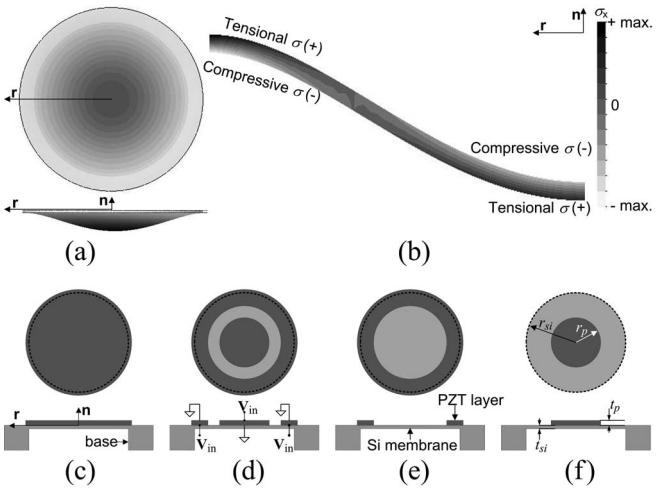


FIG. 3. Schematic design of the unit pMUT structure according to the deformation characteristics: (a) deflection of the circular plate; (b) stress σ_r in the cross section ($0 \leq r \leq r_{si}$); (c) general pMUT structure; (d) driven out of phase; (e) edge-driven; (f) centrally driven.

transducer for a given operating frequency. This improvement in mechanoacoustic efficiency means that the electroacoustic efficiency (η_{EA}) (Ref. 19) can also be enhanced.

Although the mechanical Q-factor for many materials is known, it is difficult to calculate the precise value of Q_m of a unit driver because the exact boundary conditions of the structures are difficult to determine. Therefore, Q_m of a vibration system is frequently estimated by measuring its mechanical responses at the resonance frequency in vacuum. Because Eq. (10) ultimately takes the same form as Eq. (1), $\eta_{MA,T}$ can be evaluated practically by measuring the efficiency of one component, $\eta_{MA,i}$. A detailed evaluation of this is to be continued in Sec. IV A.

Although the high mechanical resistance broadens the resonance peak band, the given difference frequency is too big for the primary frequencies to lie in a single resonance peak band. The unit driver must be driven at its resonance frequency to generate sounds with sufficiently large amplitudes. Hence, two unit drivers with two different resonance frequencies,[36] f_1 (100 kHz) and f_2 (140 kHz), are used as array components. The pMUT unit design was completed before the rest of the array was designed.

C. Unit actuator design

A PZT uni-morph was selected as the transduction mechanism because it is suitable as a resonant radiator and representative of micro-machined acoustic transducers.[15,16,20,23] The unit actuator consisted of a thin silicon plate with a PZT layer on its top surface. The radius and thickness of the circular silicon plate were selected to achieve a resonance frequency slightly above the desired primary frequency. It was important to choose the shape and the dimensions of the PZT layers to increase the diaphragm velocity for a given electrical energy and achieve a high electromechanical coupling coefficient.[15,16,37,38]

Before observing the effect of the PZT dimensions on the electromechanical coupling coefficient, the stress distribution in the cross-section [along the r-direction in Fig. 3(a)] of the silicon plate was calculated without the PZT layer using a finite element method program when the plate was deflected, as shown in Fig. 3(b). On the top surface, compressive stress occurred in the central region, but tensional stress occurred at the edge. This meant that the deposition of the PZT layer on the whole silicon plate, as shown in Fig. 3(c), may have decreased the electromechanical coupling coefficient.[37,38] Similar conclusions using top-electrode dimensions have also been reported,[39] but most research into pMUTs has not considered the effects of the PZT radius on the coupling effects in addition to its thickness. Because the PZT radius is more important in resonance-type microsystems in air (because the PZT provides relatively high damping compared to the silicon substrate), any unnecessary PZT regions had to be etched out.

Two methods to shape and drive the PZT layers for highly efficient actuation based on this physical phenomenon were devised. One method used out-of-phase voltages to drive each side of the PZT layer [Fig. 3(d)]. Although this enhances both compressive and tensional motion and seems to be the most powerful driving method, it is not compatible with large array structures because it requires complex metal line patterns. This causes practical problems such as difficulties in the PZT poling procedure or parasitic impedance[15] reduction due to the complex metal lines. Therefore, partially driving one of the two PZT regions was considered [see Figs. 3(e) and 3(f)]. The central driving method was selected for this research.

Figure 3(f) shows the basic structure of the unit radiator, which is a thin-film plate with a piezoelectric layer attached to its top surface. The ratios of the thickness and radius of the silicon plate and the PZT layer should be selected to have high coupling effects. Analytically well-derived optimal dimensional ratios using the coupling concept have been reported[38] but these ratios could not be applied in the present study because of thickness limitations in PZT fabrication and different substrate material and boundary conditions. Therefore, a finite element method was used to estimate the (k_c) effective coupling coefficient[31]

$$k_c^2 = 1 - \omega_0^2/\omega_a^2, \tag{11}$$

where ω_o and ω_a are the resonance and antiresonance frequencies, respectively. The effective coupling coefficient was estimated by obtaining the resonance and antiresonance frequencies through electrical admittance calculations using a multi-domain finite element method program. Internal damping and acoustic loading were ignored in the calculation. The radius and thickness of the silicon plate were selected to be 600 and 15 μm, respectively, for the f_2 unit driver. Few material properties of the thin-film PZT are currently known. The calculation was therefore based on the properties of bulk PZT 5H,[34] which has the same composition as the thin film (Zr/Ti=52/48) (Ref. 40) used here. Some measured properties by the manufacturer are until now listed to compare them with the bulk values in Table I.

In Fig. 4(a), the curve represents the change in the effective coupling coefficient in the f_2 unit drive with respect to the radius of the PZT layer with a constant thickness of 2.5 μm, which was the maximum consistent thickness of

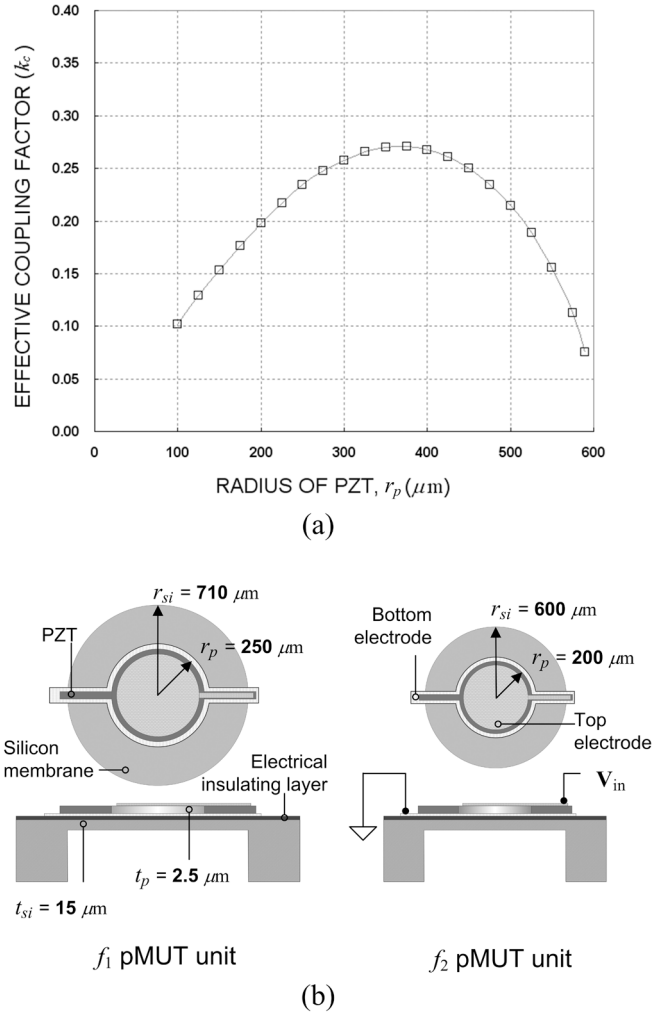


FIG. 4. (Color online) Unit actuators design: (a) effect of the PZT radius on the effective coupling factor for the f_2 pMUT unit and (b) two different resonance frequency-type pMUT units and their schematics.

PZT possible with the wafer-level sol-gel coating and etching process. The coupling coefficient increased as the radius of the PZT increased but decreased for a value of r_p greater than 350 μm. This can be explained by the stress distribution illustrated in Fig. 3(b). However, the maximum value was not used directly in the real actuator design for the input electrical impedance of the whole transducer array. Because all the pMUT units were connected in parallel to ensure in-phase operation, the total electrical capacitance, which was proportional to the top-electrode area of the whole array transducer, was too large when the optimized PZT radius was used. To work around this, the PZT radius was changed to 200 μm. Even with this design value, the effective coupling coefficient of 0.198 was larger than the whole PZT deposition on the thin-film plate [i.e., r_p=600 μm in Fig. 4(a)].

The f_1 pMUT unit was also designed using this process. The resulting two resonance frequency-type pMUT units are shown at the bottom of Fig. 4(b). The outside PZT layer ($r > r_p$) was etched out to reduce internal mechanical damping in the PZT, and only narrow PZT lines were used to connect the top and bottom electrodes to the metal lines of the array outside of the membrane [top of Fig. 4(b)].

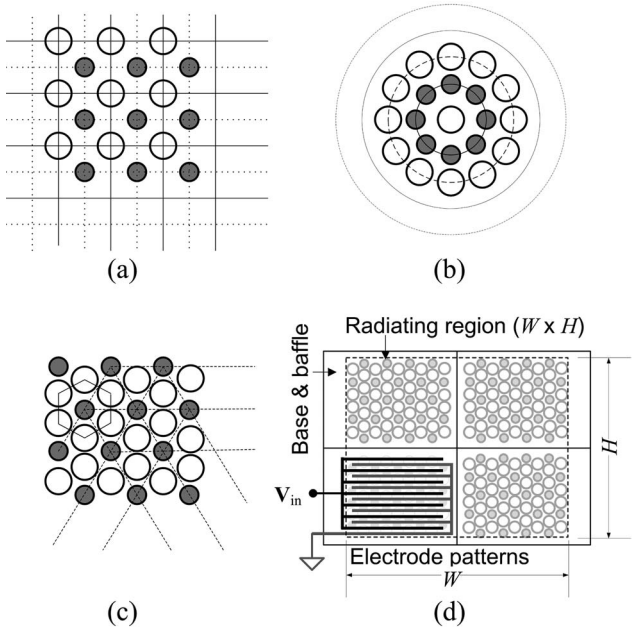
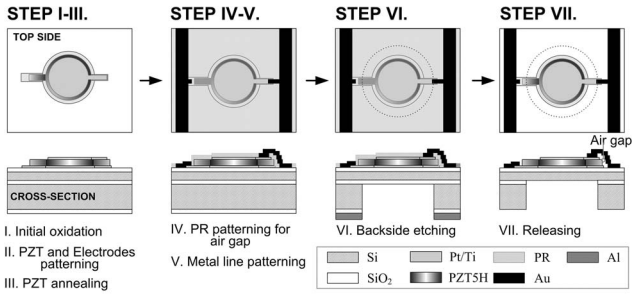


FIG. 5. Several arrangements for mixing the two types of unit radiators (white and gray circles denote unit radiators and f_1 and f_2 indicate resonance frequencies, respectively): (a) square arrangement, (b) ring-type arrangement, (c) hexagonal arrangement, and (d) design using four hexagonal sub-arrays.

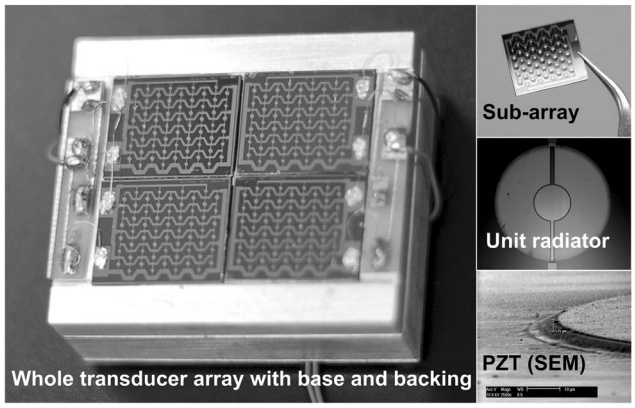
D. Transducer array design

Because each primary beam was assumed to be a collimated plane wave in the PA, a basic plane array shape was suitable. There were two design goals in creating the array. One was a good mix of the two different types of unit radiators to ensure the proper interaction between the two primary waves. This meant that the two types of unit radiators had to be repeated on the planar base.[30,36] The other requirement was to improve the radiation output power by increasing the radiation area. The efficient unit radiator did not produce enough power on its own because its active radiating area was too small. Therefore, the output power was increased by the use of an array of many small units. Several arrangement methods[30,35,36] of satisfying these requirements were possible, as shown in Fig. 5. The hexagonal arrangement shown in Fig. 5(c) was chosen for the design because the acoustic loading was expected to be more uniformly distributed by the mutual acoustic impedance.[35,41]

The resulting pMUT array was different from the assumed circular transducer with a Gaussian pressure source distribution. A bulk micro-machined array should have a passive region with zero velocity like a baffle between active units, and a square array with a similar radiation area [$W \times H = 29 \times 26$ mm$^2 \approx \pi(32/2$ mm$)^2$] was selected to increase the fabrication yield instead of using a circular shape. To overcome practical difficulties with uniformity and yield in the fabrication of a large array, the complete transducer consisted of four sub-arrays. Each sub-array had 36 f_1 units and 22 f_2 units, and the units were electrically connected in parallel to generate plane waves. Two symmetrical comb-shaped electrical metal lines were used to achieve a similar electrical metal line resistance for connecting the units in parallel. The final transducer array is shown in Fig. 5(d). The total capaci-



(a)



(b)

FIG. 6. Fabrication of the micro-machined source transducer: (a) schematic view of the fabrication process and (b) assembled micro-machined source transducer.

tance of the PZT in the designed array was 173.2 nF.

III. FABRICATION

Figure 6(a)shows the basic fabrication process. A silicon on insulator wafer was used to form the uniform thin film for the array transducer. The device layer, the insulating SiO_2 layer, and the handling layer of the silicon on insulator wafer were 15, 1, and 500 μm thick, respectively. Initial thermal oxidation was applied to the electrical insulating layer between the device silicon layer and the bottom electrode, and the thickness of the SiO_2 was 6000 Å (step I). Pt and Ti were deposited using sputtering on the top of the device layer with thicknesses of 2500 and 330 Å, respectively, to form the bottom electrode. The 2.5-μm-thick PZT layer was coated in the sol-gel process using PZT solution (Zr/Ti=52/48, morphotropic composition; its properties have been described previously[40]). A 2500-Å-thick Pt layer was deposited by sputtering on the PZT layer for the top electrode. After depositing those materials, the electrodes and the PZT layer were patterned by dry etching (step II). A short period of rapid thermal annealing at 650 °C ensured that the PZT layer recovered from the etching damage (step III). The air-gapped pad connection method was used to connect the electrical wire to the top electrode after isolating the top and bottom electrodes.[42] The air-gapped pad connection required three main procedures. First, the sacrificial photoresist (PR) layer was added and patterned onto the space between the top and bottom electrodes (step IV). On the patterned PR, a 3300-Å-thick layer of gold was deposited and patterned by dry etching for the connection (step V); the PR layer was removed in step VII. The backside was etched with an induced coupled-plasma reactive ion etcher using aluminum window patterns to create a thin-film structure (step VI). The complete array transducer was fabricated after releasing the SiO_2 layer between the device and the handing layers (step VII). The air-gapped connection was built after removing the PR. As shown in Fig. 6(b), the fabricated micro-machined transducer arrays were installed on the base, which consisted of a baffle and absorptive backing material. The whole micro-machined structure size was about 35×33 mm^2, including the edge handling portion. Two sets of micro-machined source transducers were constructed and designated as source transducers I and II.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Before the experiments to generate the DFW with the PA were performed, the dynamic characteristics related to primary wave generation with the pMUT unit samples and the pMUT array were evaluated. The pMUT unit samples were fabricated simultaneously during the pMUT array fabrication. Radiation characteristics of the difference wave were observed to confirm the PA in air. Figure 7 shows the experimental setups for evaluating the vibration and acoustic radiation characteristics. Acoustic measurements were performed in a semi-anechoic chamber 3 m wide, 3 m long, and 2 m high. The ambient acoustic noise in the chamber was below 15 dB re 20 μPa above 150 Hz, and the ambient temperature was assumed to be 20 °C. The transducer and microphone were installed at a height of 1.2 m to reduce the effects of floor reflection during the experiments.

A. Evaluations of unit pMUT performance

The mechanoacoustic efficiency [$\eta_{MA,i}$ in Eq. (1)] of the pMUT unit sample was estimated first. The internal damping of underwater transducers is often measured in air because the acoustic loading in air is negligible compared to that in water.[31] However, the acoustic loading in a micro-machined ultrasonic/acoustic transducer in air is expected to be relatively large compared to the internal damping because of the thin-film plate. To estimate R_{mi} and R_{ri}, a vacuum chamber with a minimum pressure of approximately 0.5 kPa was designed for the velocity response test [Fig. 7(b)]. The vibration velocity amplitude, according to the sweep sinusoidal signal, was measured at the center of the sample, as shown in Fig. 8(a), using a laser Doppler vibrometer [input line **A** and output line **C** in Fig. 7(a)].

The mechanical Q-factor of the f_1 unit sample (f_2 unit sample) was 27 (65) and 275 (378) in air and in vacuum, respectively. Assuming no viscous damping and ignoring the radiation mass [i.e., $X_{ri}/\omega\approx0$ in Fig. 2(b)], these large differences in the Q-factor simply mean that symmetrical acoustical loading at both membrane sides (i.e., top and bottom), $R_{\text{acoustical}}=2R_{ri}\approx\omega_0 m_c(1/Q_{\text{air}}-1/Q_{\text{vac}})$, was very large compared to the internal damping, $R_{mi}\approx\omega_0 m_e(1/Q_{\text{vac}})$,[31,43] where m_e and ω_0 are the equivalent mass of the plate including the PZT layer and the resonance frequency, respectively. In other words, almost all of the mechanical energy was consumed by the acoustical loading. Considering only top

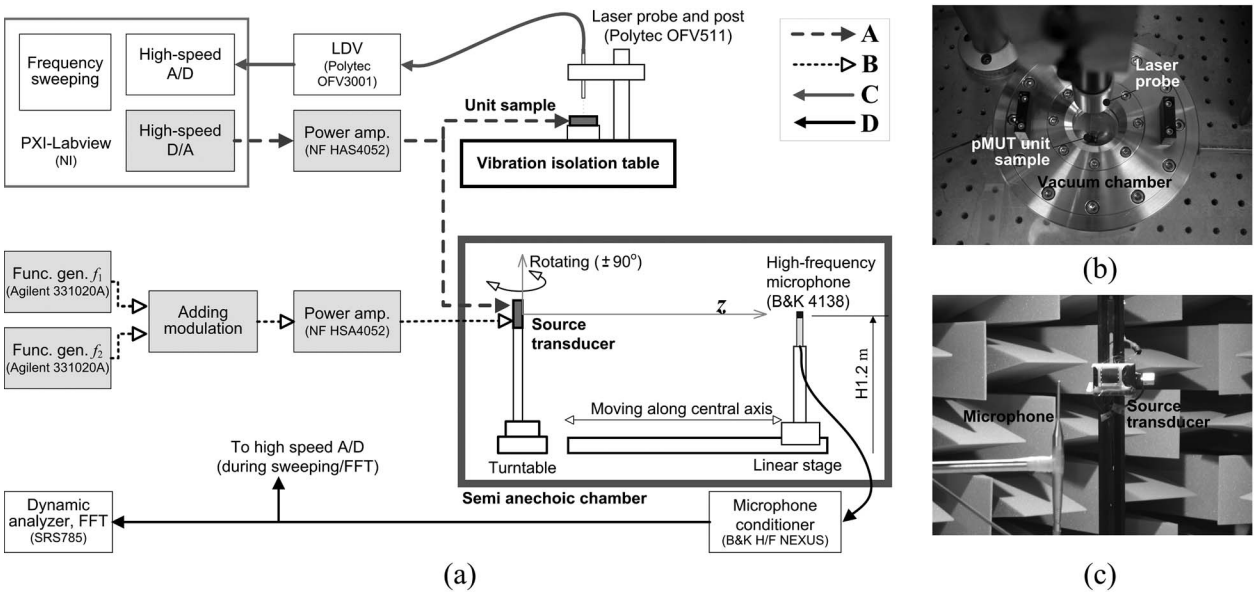


FIG. 7. Experimental setup to evaluate dynamic characteristics and PA generation. **A** is the input signal line, **B** is the PA input signal line, **C** is the laser Doppler vibrometer output signal line, and **D** is the microphone output signal line. (a) Experimental schematics; (b) the vacuum chamber and specimen; and (c) setup in an anechoic chamber.

side radiation [i.e., $R_{ri} \approx 0.5 \times \omega_0 m_e (1/Q_{air} - 1/Q_{vac})$], the mechanoacoustic efficiency was estimated using Eq. (1) to be about 82.1% and 70.7% at the resonance frequency in air for the f_1 and f_2 unit samples, respectively. This efficiency is very high compared to conventional transducers and indicates the micro-machined source transducer's potential for efficient radiation in air.[9,19]

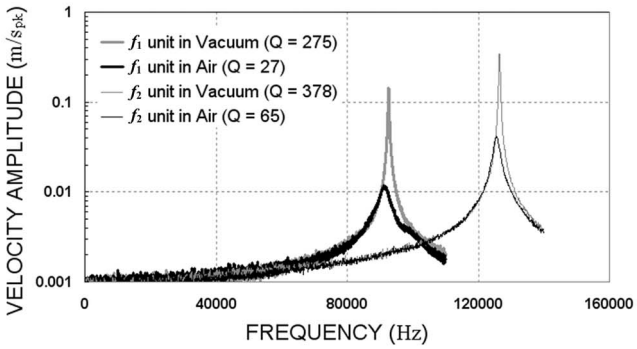
The velocity amplitude of the f_1 unit sample (f_2 unit sample) at resonance in air increased in linear proportion to the input voltage level before 5 V_{pk} (6.5 V_{pk}) and saturated around 10 V_{pk}, as shown in Fig. 8(b). The velocity amplitude response of the f_1 unit sample (f_2 unit sample) up to 5 V_{pk} (6.5 V_{pk}) was approximately 0.42 m/s_{pk} per 1 V_{pk} (0.48 m/s_{pk} per 1 V_{pk}), which is a fairly good dynamic response;[23] a maximum velocity of 1.88 m/s_{pk} (3.12 m/s_{pk}) was achieved. The corresponding sound pressure at the center of the thin-film plate for the maximum velocity of f_2 was calculated by Rayleigh integration[19] with the edge-clamped boundary condition[44] as 915.2 Pa_{pk}. This indicates the potential to use the designed micro-machined transducer as a high-intensity source transducer. However, the pressure distribution of the plate and the baffle region of the array were also taken into account in evaluating its performance.

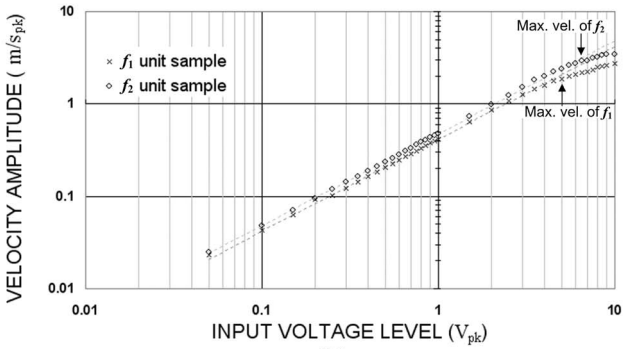
The input electrical admittance[19] of the f_2 unit sample was measured, and the result is shown in Fig. 8(c). Simulated ideal admittance curves are also plotted in this figure, and some parasitic impedance components such as metal line resistance and parasitic impedance through the semiconducting substrate,[13] which are by-products of micro-machining, were detected. This decreased the ideal electroacoustic efficiency, which was around 68% (f_2 unit sample) at resonance, ignoring the parasitic components.

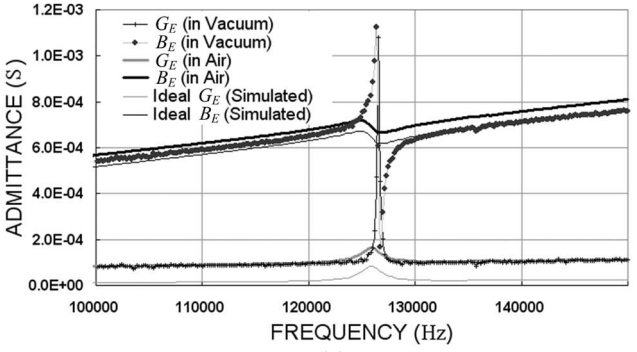
B. Evaluation of the fabricated transducer array

To test the frequency response of the whole pMUT array, the output SPL at a distance of 1 m from the transducer was measured for an input of 3 V_{pk} and a dc bias of −20 V using the test setup illustrated in Fig. 7(a) (input line **A** and output line **D**). The applied dc bias voltage prevented depolarization of the PZT layers, and therefore similar transducer conditions were maintained during the experiments. To use the high-frequency pressure field microphone (B&K 4138 with a high-frequency conditioning amplifier) in a free-field measurement, the microphone was installed perpendicular to the central axis of the transducer [Fig. 7(c)].[45] Additional post-processing was needed in this configuration to acquire an actual response because high-frequency sensitivity was largely reduced. In addition to the direct microphone output signal (measured response), the post-processed data (corrected response) considering free-field correction[19,45] for the 90° incident angle and actuator response of the microphone itself are plotted in Fig. 9. The following analysis and experiments regarding PA were based on the measured response data because the post-processing procedure had uncertainty issues caused by the manual correction process that depended on microphone calibration charts. However, the correction values as a function of the frequency used in this work are also supplied at the bottom of Fig. 9.

Two major peak regions appeared around the designed resonance frequencies, but the peaks were not as clear as those of the velocity responses of unit samples shown in Fig. 8(a). The difference in peak levels was mainly due to the difference in the overall radiation areas of the f_1 units and f_2 units. In case of source transducer I (source transducer II), two major peaks occurred at 94.6 kHz(103 kHz) and 131 kHz(136.5 kHz). Some minor peaks at 101.5 kHz (95.8 kHz) and 134.5 kHz were caused by the nonuniformity of the resonance frequencies in each unit radiator. The lack of clarity in the output SPL response was due to the footing effect,[46] an over-etching phenomenon that occurs in deep-reactive ion etcher processes during the backside bulk etching (depth=500 μm) for thin-film formation. It increases







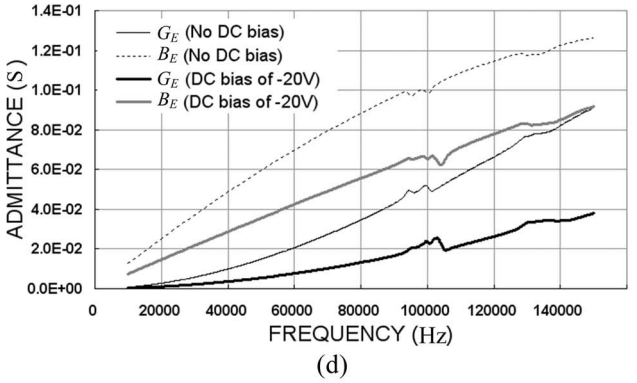


FIG. 8. Dynamic characteristics of unit actuator samples: (a) velocity response in air and in a vacuum, (b) velocity amplitude in the membrane center according to the input voltage level (f_1 pMUT unit sample), (c) the input electrical admittance (f_2 pMUT unit sample), and (d) the input electrical admittance (source transducer II).

the plate radius of some units at the edge of the wafer, causing an irregular decrease in resonance frequency. The resonance frequencies of the test samples in Figs. 8(a)–8(c) also decreased for the same reason. Methods to reduce this effect are under investigation. This non-uniformity in resonance

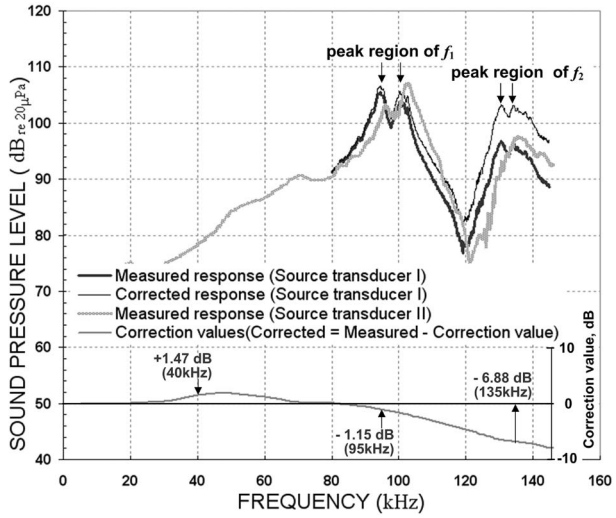


FIG. 9. Output SPL of the transducer as a function of frequency ($z=1$ m in front of transducer, $V_{in}=3$ V_{pk}, and -20 V_{dc} bias) and microphone correction value.

frequencies directly decreased the SPL of the primary waves. The input electrical admittance of source transducer II was measured as shown in Fig. 8(d). However, the curves were strongly distorted by complex parasitic impedance components and the non-uniformity of the resonance frequencies. This was also related to the dc bias voltage. Further analysis of the transducer array gave meaningless results because a solid understanding of all the relations was not obtained.

C. DFW generation

The experiments on DFW generation were conducted with continuous input signals. The experimental setup is shown in Fig. 7(a) (input line **B** and output line **D**). Two dc-biased sinusoidal signals with primary frequencies were generated by two function generators in additive modulation and applied to the micro-machined source transducer after they passed through a power amplifier. Considering the output pressure responses, f_1 and f_2 in the PA experiment were changed to maintain $f_d \cong 40$ kHz. The pressure amplitude of each primary wave was controlled by adjusting the input signal gains. The final input driving signal was

$$\mathbf{V}_{\text{in}}(t) = \mathbf{V}_M(t) + V_{\text{dc}}, \tag{12}$$

where $\mathbf{V}_M(t) = V_1 \cdot \sin(2\pi f_1 t) + V_2 \cdot \sin(2\pi f_2 t + \varphi)$, $V_2/V_1 = \chi$ considering the measured response in Fig. 9) is the modulated driving signal and V_{dc} is the dc bias voltage to polarize the PZT layer. Detailed operating conditions of each source transducer are listed in Table II. For the directional pattern measurements, the high-frequency microphone was installed 0.8 m from source transducer I on the turntable, and the SPL was measured by rotating the transducer. The propagation curves were measured by moving the microphone on a linear rail with source transducer II. Data from one primary wave of f_1 were also plotted on the related graphs. The f_2 data were omitted because of limitations in the fast Fourier transform equipment, but the SPL of f_2 at $z=1$ m was checked with an oscilloscope at the beginning of the experiments.

TABLE II. Operating conditions for source transducers I and II.

	f_1	f_2	f_d	χ $(=V_1/V_2)$	V_{dc}	Related PA experiments in Chapter IV. Section C
Source transducer I	95.0	134.5	39.5	1.5	−20	2. Beam patterns
Source transducer II	95.0	135.0	40.0	1.25	−20	3. Propagation curves and 4. Maximum performance
Units	kHz	kHz	kHz	—	V	

Difficulties have been often reported in estimating the actual DFW signal of a PA through eliminating various spurious difference signal sources caused by the receiving and transmitting systems.[6,30,36] However, the reduced sensitivity in high-frequency ranges using the perpendicular microphone configuration gave an opportunity to reduce the sources in the receiving system.[47] Compared to a normal incident microphone configuration (+7-dB increment in sensitivity of the microphone used), about −14 dB of the f_2 (135 kHz) primary wave signal could be compressed by this perpendicular installation. Therefore, it worked as a low-pass filter to compress one (f_2) of the primary components. Furthermore, the small downshift ratio between primary and secondary waves $(f_m/f_d \approx 2.9)$ was also expected to reduce the difficulties with spurious signals in this work.[6] Parasitic radiation[6] caused by the transmitting system was expected to be negligible because the two different resonance-type pMUT units were operated nearly independently with a small input voltage level.

1. Models for theoretical references

Appropriate theoretical reference data are required for comparison. However, the exact predictions of the DFW for the fabricated transducer are difficult to discern because of the complex source configurations and the short measurement range (remaining in the near-field of the PA). Therefore, simple ideal cases were selected for prediction, and analytic calculations were performed to generate trends of the DFW for the given primary fields, as follows:

- *Model I*. Rectangular array consisting of ideal circular piston elements to predict primary beam patterns by summing the far-field wave of each element assuming in-phase motion,[19] where each element location and size is matched with the actual transducer; element velocity is a normal value.
- *Model II*. PA model with an axisymmetric Gaussian source,[30] which was used previously in the transducer design to predict the DFW field in the far-field of the source transducer.
- *Model III*. Fenlon solution for collimated infinite plane waves[48] to predict the DFW field near the source transducer.

It is necessary to compare the differences between the actual micro-machined transducer and these theoretical models before addressing the PA experimental results. The main difference in the source pressure distribution between model I and the actual micro-machined transducer is in each element. The actual element is the edge-clamped circular disk, which is slightly deflected to the rear due to the residual stress. In fact, the phases are not identical due to nonuniformity in the resonance frequencies. As discussed in the section dealing with array design, the source distribution on the fabricated transducer array also differs from the axisymmetric Gaussian source of model II and is closer to a rectangular uniform source. The source distributions of these models are plotted using Cartesian coordinates in Fig. 10(a).

2. Beam patterns

The beam pattern of the f_1 primary wave was measured [on the z-x plane in Fig. 10(a)] during the PA experiment. Figure 11(a) shows the measured and calculated beam patterns (ideal transducer array (i.e., model I with f_1 elements and ideal Gaussian source, i.e., model II). The measured beam pattern was similar to that of the estimated ideal transducer array without level differences (less than 5 dB) in some side lobes. Slight unsymmetrical patterns in the measured data were caused by a tilting error in the installation of the transducer. Beam patterns of f_2 were also estimated by the same methods, as shown in Fig. 11(b). Except for the complex side lobes, the major lobe pattern of the transducer array was quite similar to that of the ideal Gaussian source, meaning that a difference frequency beam pattern compari-

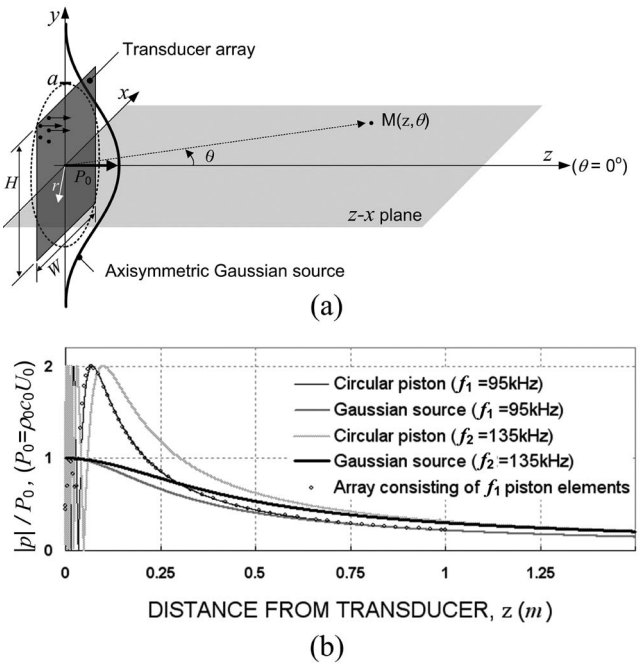


FIG. 10. Source distributions on radiators and calculated pressure amplitude on the z-axis (Refs. 19 and 20): (a) source distributions of the Gaussian and the transducer array; (b) axial pressure amplitudes of the circular piston, the Gaussian source, and the transducer array.

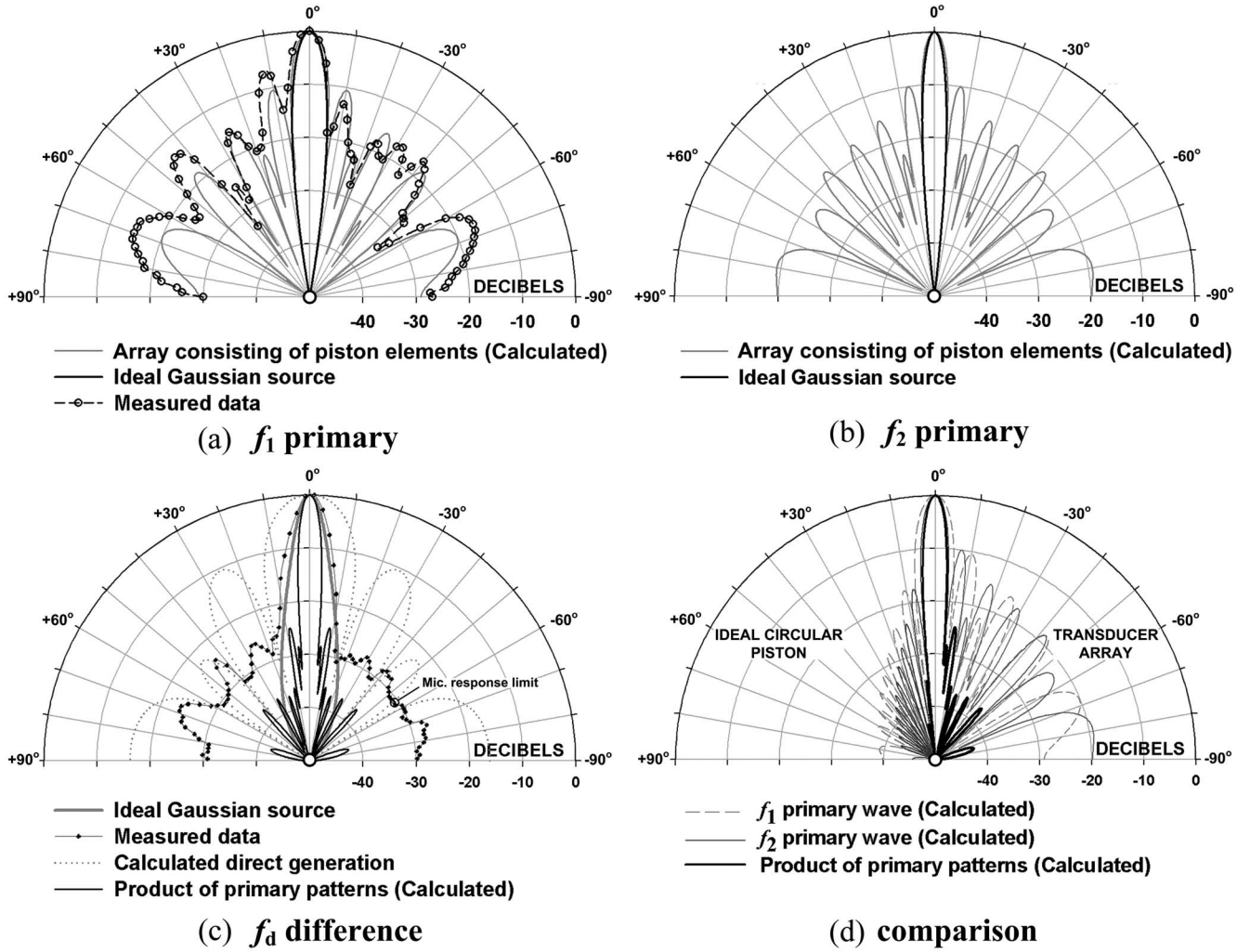


FIG. 11. Beam patterns of the primary and DFWs (z=0.8 m in front of transducer); (a) primary wave, f_1=95 kHz, (b) primary wave, f_2=134.5 kHz, (c) difference wave, f_d=39.5 kHz and product of calculated primary patterns, and (d) calculated primary waves and product pattern of two different sources.

son between the two cases was possible in the discussion of PA generation based on the micro-machined source transducer.

The measured difference frequency beam pattern (solid gray line with black dots) is compared to the theoretical result (solid gray line, model II, using Eq. 6.20 from Ref. 30) in Fig. 11(c). The beam patterns of the major lobes were quite similar. The measured HPBW of the DFW was around 5°, as expected from the design. Even the levels were very small compared to the major lobe (at least −23.8 dB difference), and some minor lobes were observed. To identify the main source of those minor lobes, the product of the simulated primary beam patterns is also plotted in Fig. 11(c). Close correspondence was observed between the angle of the minor lobes and that of the partially collimated primary beams. However, whether the minor lobes of the DFW were caused by the PA or by spurious signals as discussed in Ref. 6 was unclear. Therefore, we measured propagation curves before concluding whether the measured beam pattern was really that of the DFW generated by the PA.[6]

3. Propagation curves

Figure 12 shows the measured propagation curves of the DFW (square dots) and the primary wave (plus dots) for a modulated driving voltage level, $|\mathbf{V}_M(t)|$, of 3 V_{pk}. The measurement for distances greater than 1 m to observe the far-field characteristics of DFW could not be performed due to the limitations of the anechoic chamber. The simulated results for the ideal Gaussian source (model II, using Eq. 6.18 in Ref. 30) are also plotted in this figure for comparison.

Some near-field differences in primary waves occurred because the pressure distribution in the real transducer array (plus dots) was closer to being a uniform source (black dotted line, labeling **c**, $2a$=32 mm) rather than the assumed Gaussian source (heavy black line, $2a$=32 mm). The circular piston was introduced to represent the axial pattern of a uniform source, such as a local maximum, but more realistic behavior near the transducer was that of the ideal transducer array (thin black line, model I). To compare trends in PA generation, the source pressure amplitude (P_1 and P_2=P_0) of two primary waves in the calculation was set to half the local maximum pressure amplitude ($P_{max} \cong 2P_{1\ piston}$=$2P_0$) of the measured primary wave data for convenience.

DFW generation trends of the measurement data (plus dots for f_1 and square dots for f_d) were basically similar to the theoretical results (heavy black line for f_1 and heavy gray line for f_d). However, DFW generation near the transducer of

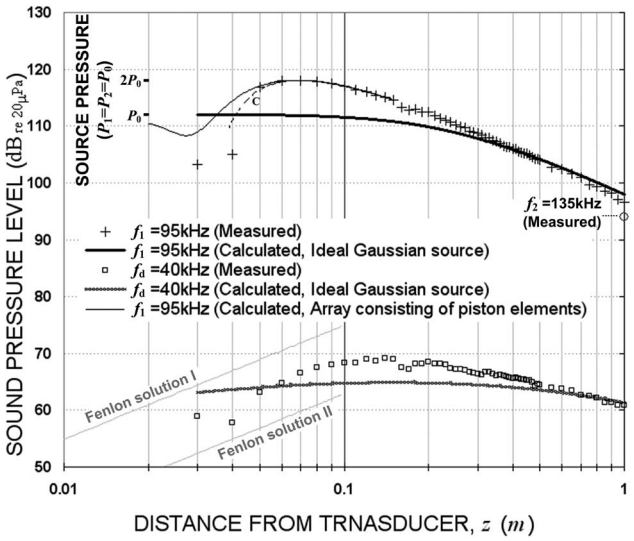


FIG. 12. Propagation curves.

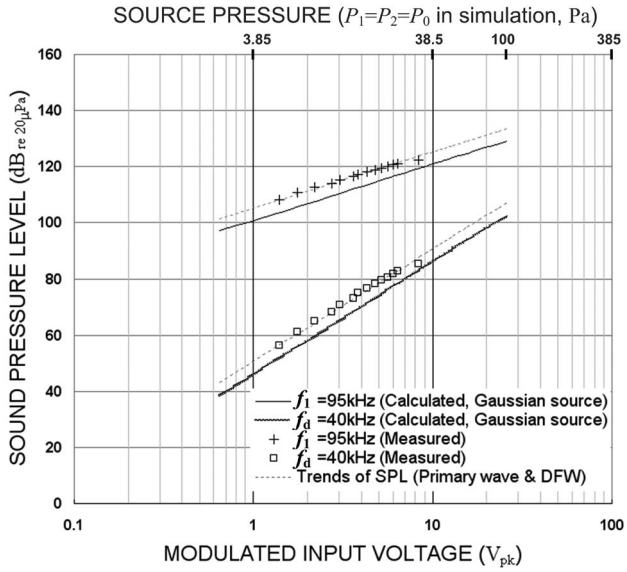


FIG. 13. SPL of primary and secondary waves as a function of input modulated voltage level at z=0.15 m from the transducer.

a Gaussian source is known to be different from a uniform source[49] like the fabricated transducer. Therefore, Fenlon solutions[48] (model III), which describe DFW generation of collimated plane waves, were calculated and are illustrated in the propagation curve to compare the characteristics near the transducer ($z \leqslant 0.1$ m; P_1 and $P_2=2P_0$ for Fenlon solution I and P_1 and $P_2=P_0$ for Fenlon solution II). Unlike the basic assumption in Fenlon solution, the measured pressure amplitude of the primary wave near the transducer was not constant. Therefore, the two solutions were used as maximum and minimum limits. Even though a slight oscillation (at z =0.03 and 0.04 m) in the measured propagation curve of DFW was detected, the increased pattern near the transducer agreed with the theoretical trend. If the spurious microphone signal had a dominant effect on the measured data of the DFW, then a rapid pressure drop (close to −12 dB per doubling distance; see Refs. 6 and 30) would be expected where spherical spreading of primary waves takes place. Therefore, based on the results in Figs. 11 and 12, it appears that the DFW of the PA was successfully generated using the micromachined source transducer.

Unexpected measured data of the primary wave at z =0.03 and 0.04 m should be noted. Their SPLs were much lower (<10 dB) than that of the ideal transducer array (thin black line, model I). The main reason for this is not yet clear due to experimental considerations, such as phase differences in array elements or short measurement distance near the transducer.

With regard to the minor lobes of the DFW in Fig. 11(c), the beam patterns in Fig. 11(d) were calculated assuming a general circular piston transducer to compare them to those of the transducer array. The minor lobes are generally treated as indicators of spurious signals in a measurement or direct sound generation, but in this case the product of the simulated primary beam patterns showed that those minor lobes could also be DFWs generated by the PA. That is, partially collimated side lobes with significant SPL may generate those minor lobes of the DFW. However, to improve the practicality of the source transducer, those minor lobes should be compressed through other considerations in the array design.

4. Maximum performance

To evaluate the maximum performance of the micromachined source transducer, the SPLs of the DFW and primary wave were measured at 0.15 m, where the SPL of the DFW peaks was in Fig. 12, by increasing the modulated input voltage level. Theoretically, the output sound pressure of the DFW is proportional to the value of the two pressure amplitudes of the primary waves multiplied together, assuming no saturation effects,[1] and thus the difference between the SPL of the DFW and primary waves should decrease as the input voltage level increases. Figure 13 shows that each pair of measured and calculated data agreed. The measured maximum SPLs of the difference frequency and primary waves were about 85.4 and 122.5 dB, respectively, for an applied modulated driving voltage level of 8.3 V_{pk} ($V_1 \cong 3.69$ V_{pk} and $V_2 \cong 4.62$ V_{pk}). Based on the trends in Fig. 13, a higher SPL for the DFW would be expected if a higher input voltage could be applied. A higher input voltage could not be used because of thermal problems in the pMUT array; this heat was likely caused by parasitic impedance in the micro-system. Related issues are often reported,[15] but unfortunately, this problem is more common in high-frequency and high-powered micro-systems like the device developed here. A more detailed analysis to resolve this problem is currently in progress.

5. Equivalent source pressure in calculation

The source pressure amplitudes (P_1 and P_2, denoted together as $P_{1,2}$) of the Gaussian source for the PA calculations were defined simply in Sec. IV C 3. Before discussing the method used to define them and its validity, it is necessary to introduce the analytical solutions of Novikov *et al.*[30] that

were used in the DFW field calculations (Ref. 30 is not generally available). The axisymmetric Gaussian source distribution is

$$q_{1,2}(r,z=0) = P_{1,2}\exp(-r^2/a^2), \tag{13}$$

where a and $P_{1,2}$ are the effective radius and the peak source pressure, respectively. Before addressing the solutions, the reduced coordinates (z_a, θ_W) and the reduced amplitude (P_n) transformed from the physical coordinates (z, θ) and sound pressure amplitude of the DFW $(P\_)$ are defined as[30]

$$z_a = z/l_a, \tag{14}$$

$$(\theta_W)^2 = \theta^2 K l_a/2, \tag{15}$$

$$P_n = P\_8c_0^4\rho_0 \exp(z/L_a)/\beta(2\pi f_d)^2 P_1 P_2 a^2, \tag{16}$$

where $l_a = 2/(\alpha_1+\alpha_2-\alpha\_)$, $\alpha_{1,2}$ is the absorption coefficient at the primary frequency, $\alpha\_$ is the absorption coefficient at the difference frequency, $L_a = 1/\alpha\_$, K is the wave number of the difference frequency, and the physical properties ρ_0, c_0, and β are the ambient density ($\rho_0 = 1.21$ kg/m^3 at 20 °C in air), the small signal sound speed ($c_0 = 343$ m/s), and the coefficient of nonlinearity ($\beta = 1.2$), respectively.

Assuming that the ratio f_d/f_m is small, the axial pressure amplitude and the beam pattern of the DFW can be calculated numerically by the following equations:[30]

$$P_n(z_a,0) = i(i+Bz_a)^{-1}\exp\frac{d-iz_a}{i+Bz_a} \times\left[Ei\left(\frac{Bz_a+d}{i+Bz_a}\right) - Ei\left(\frac{d-iz_a}{i+Bz_a}\right)\right], \tag{17}$$

$$P_n(z_a,\theta_W) = i\int_0^{z_a}\exp\left(-y-\frac{(\theta_W)^2 z_a^2(1+iBy)}{i(y-z_a)+yz_aB}\right) \times[i(y-z_a)+yz_aB]^{-1}dy, \tag{18}$$

where $Ei(x) = \int_{-\infty}^{x}(e^y/x)dy$ is the exponential integral, and the parameters B and d are defined as follows:[30]

$$B = L_d l_a/l_{d1}l_{d2}, \tag{19}$$

$$d = L_d/l_a, \tag{20}$$

where L_d is the Rayleigh distance of the difference wave. Equation (18) is valid only at $z \gg l_{d1,d2}$, where $l_{d1,d2}$ is the Rayleigh distance of the primary wave.

The main issue in the axial pressure amplitude calculation with Eqs. (16) and (17) is how to properly define the source pressure $P_{1,2}$ of the Gaussian source from the measured primary fields. It is difficult to define the value directly with the transducer array because the Gaussian source is an axisymmetric source with an ideally infinite radial distribution. The source distributions on the fabricated transducer are closer to the finitely distributed uniform source (especially a rectangular piston with $W \times H \approx \pi a^2$, where $H/W \approx 0.9$) rather than the Gaussian source. A circular piston, the radius of which is the same as the effective radius a, is considered for convenience. Its source distribution is axisymmetric like the Gaussian source, as well as uniform like the actual transducer array. Therefore, it seems to be a link between the two different sources.

If the velocity amplitude of the circular piston (let U_p) is the same as the peak velocity of the Gaussian source (let U_0), their volume velocities are identical to each other. Moreover, in the case of $z \gg a$, they show practically identical axial pressure behaviors.[30] Figure 10(b) shows their axial pressure amplitudes, calculated for the two primary frequencies (f_1 and f_2), when the two velocity amplitudes are the same (i.e., $U_p = U_0$). The source pressure amplitude P_0 of the Gaussian source is half of the local maximum value (let $P_{\max}$; i.e., $P_{\max} = 2P_0$) of the circular piston. The axial pressure pattern (circular dots) of the ideal array (model I) was also calculated only with f_1 elements of unknown velocity amplitude (chosen arbitrarily, and assuming f_2 elements had no significant effect on it). Its level was adjusted by dividing by a constant factor to make the local maximum value of 2. The axial behavior of the ideal array model showed characteristics that were quite similar to those of the uniform piston source, except for the axial patterns near the transducer. In addition, the collimated beam region (i.e., the interaction region distributed along the z-axis) was also approximately similar to that of the uniform source. Therefore, the equivalent source pressure of the ideal transducer array could also be defined in practice considering the relation between the circular piston and the Gaussian source. That is, the existing local maximum value (in the axial pattern of the transducer array, model I) was twice the source pressure.

The measured axial patterns of the f_1 primary wave were also similar to those of the circular piston, including the local maximum point ($z = 0.07$ m in Fig. 12), as for the ideal transducer array. Therefore, the equivalent source pressure $P_{1,2}$ in the calculation model could be defined. Half the local maximum pressure value ($P_{\max}$) of the measured f_1 primary wave was selected as the source pressure in the calculation of the solution (i.e., $P_1 = P_{\max}/2$). This source pressure was used in the calculation of the DFW to predict the far-field of the transducer. In the calculations to determine the DFW fields in Fig. 13, a similar method was used to define the source pressure of the primary wave. That is, the measured primary wave data were assumed to be the pressure amplitude of the ideal circular piston at the same position.

V. CONCLUSION

A MUT and array were designed as a new source transducer in air based on the proper operating condition of the PA. The experiments with the unit samples showed the potential for developing a highly efficient, thin, small primary wave projector. The mechanoacoustic efficiencies of the f_1 and f_2 units were 82.1% and 70.7%, respectively, and the ideal electroacoustic efficiency ignoring parasitic impedance was more than 68%. The MUT also had a noticeable vibration velocity response, which indicated its potential for use as a micro-machined source transducer.

A DFW was generated by the MUT array designed here. The measured beam patterns and the SPL of the DFW were quantitatively similar to the theoretical results obtained from

the ideal models, indicating that the DFW was successfully generated by the PA. The beamwidth was 5°, and the maximum output SPLs of the primary and secondary waves were 122.5 and 85.4 dB, respectively. Even better performance is expected after certain practical problems are addressed.

Two practical issues, those of fabrication uniformity and parasitic impedance, are often considerable in a large array transducer with a high resonance frequency. Detailed analysis and adjustments to the fabrication conditions to reduce these problems are in progress, and results are expected soon.

The transducer can be used as a highly directional ultrasonic transducer for enhancing the spatial resolution in ranging measurements in air, just as in water, because it has a very narrow beamwidth that cannot be achieved using a conventional direct radiation method with the same transducer dimension and operating frequency.

With more theoretical and experimental research, devices for other practical applications will be possible. These could include highly efficient transducers for small or thin directional loudspeakers in mobile appliances. The basis for some of this development has been provided in this paper, and the additional work required to improve the potential of the transducer is already in progress.

ACKNOWLEDGMENTS

The authors would like to thank Mr. S. Cho and Mr. Y. Jae in POSTECH for their help in some numerical calculations and for helpful discussions. We have cooperated with Micro-To-Nano Inc. and INOSTEC Inc. to fabricate the device described here. This work was supported in part by the IT R&D program of MIC/ITA (Grant No. 2005-S-111-02), and in part by a grant to MEMS Research Center for National Defense funded by Defense Acquisition Program Administration.

[1]P. J. Westervelt, "Parametric acoustic array," J. Acoust. Soc. Am. **35**, 535–537 (1963).
[2]H. O. Berktay, "Possible exploitation of nonlinear acoustics in underwater transmitting applications," J. Sound Vib. **2**, 435–461 (1965).
[3]J. L. S. Bellin and R. T. Bayer, "Experimental investigation of an endfire array," J. Acoust. Soc. Am. **34**, 1051–1054 (1962).
[4]M. E. Zakharia, S. Hibral, C. Pollet, and E. Rigaud, "Parametric synthetic interferometric sonar for bottom and subbottom imaging and mapping," J. Acoust. Soc. Am. **116**, 2577 (2004).
[5]R. F. W. Coates, M. Zheng, and L. Wang, "BASS 330 PARACOM: A model underwater parametric communication system," IEEE J. Ocean. Eng. **21**, 225–232 (1996).
[6]M. B. Bennett and D. T. Blackstock, "Parametric array in air," J. Acoust. Soc. Am. **57**, 562–568 (1975).
[7]M. Yoneyama, J.-Ichiroh Fujimoto, Y. Kawamo, and S. Sasabe, "The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design," J. Acoust. Soc. Am. **73**, 1532–1536 (1983).
[8]A. R. Selfridge and P. Khuri-Yakub, "Piezoelectric film sonic emitter," U.S. Patent No. 6,606,389 B1 (12 August 2003).
[9]J. A. Gallego-Juarez, G. Rodriguez-Corral, and L. Gaete-Garreton, "An ultrasonic transducer for high power applications in gases," Ultrasonics **16**, 267–271 (1978).
[10]K. Naugolnykh and L. Ostrovsky, *Nonlinear Wave Processes in Acoustics* (Cambridge, New York, 1998), pp. 1–10.
[11]R. Heydt, R. Pelrine, J. Joseph, J. Eckerle, and R. Kornbluh, "Acoustical performance of an electrostrictive polymer film loudspeaker," J. Acoust. Soc. Am. **107**, 833–839 (2000).
[12]K. W. Cheung, P. P. Tong, and C. D. Thomas, "Directional wireless communication systems," U.S. Patent No. 7,269,452 (11 September 2007).
[13]W. Moon and H. Lee, "Ultrasonic ranging system and method thereof in air by using parametric array," U.S. Patent No. 7,196,970 (27 March 2007).
[14]C. Lee, J. Chang, J. Park, and Y. Kim, "Personal sound system design for mobile phone, monitor, and television set: Feasibility study," J. Acoust. Soc. Am. **122**, 3353(A) (2007).
[15]P. Muralt and J. Baborowski, "Micromachined ultrasonic transducers and acoustic sensors based on piezoelectric thin films," J. Electroceram. **12**, 101–108 (2004).
[16]G. Perçin, A. Atalar, F. L. Degertekin, and B. T. Khuri-Yakub, "Micromachined two-dimensional array piezoelectrically actuated transducers," Appl. Phys. Lett. **72**, 1397–1399 (1998).
[17]D. A. Hutchins, J. S. McIntosh, A. Neild, D. R. Billson, and R. A. Noble, "Radiated fields of capacitive micromachined ultrasonic transducers in air," J. Acoust. Soc. Am. **114**, 1435–1449 (2003).
[18]H. T. Soh, I. Ladabaum, A. Atalar, C. F. Quate, and B. T. Khuri-Yakub, "Silicon micromachined ultrasonic immersion transducers," Appl. Phys. Lett. **69**, 3674–3676 (1996).
[19]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 2000), pp. 107–109, 179–209, and 390–422.
[20]F. Akasheh, T. Myers, J. D. Fraser, S. Bose, and A. Bandyopadhyay, "Development of piezoelectric micromachined ultrasonic transducers," Sens. Actuators, A **111**, 275–287 (2004).
[21]P. C. Eccardt and K. Neiderer, "Micromachined ultrasonic transducers with improved coupling factors from a CMOS compatible process," Ultrasonics **38**, 744–780 (2000).
[22]W. Moon and I. J. Busch-Vishniac, "Modeling of piezoelectric ceramic vibrators including thermal effects. Part IV. Development and experimental evaluation of a bond graph model of the thickness vibrator," J. Acoust. Soc. Am. **101**, 1408–1429 (1997).
[23]Z. Wang, W. Zhu, J. Miao, H. Zhu, C. Chao, and O. K. Tan, "Micromachined thick film piezoelectric ultrasonic transducer array," Sens. Actuators, A **133–131**, 485–490 (2006).
[24]H. Lee and W. Moon, "Microelectromechanical-systems-based parametric transmitting array in air—Application to the ultrasonic ranging transducer with high directionality," J. Acoust. Soc. Am. **119**, 3379 (2006).
[25]H. Lee, D. Kang, and W. Moon, "Microelectromechanical-systems-based parametric transmitting array in air—Application to the ultrasonic ranging transducer with high directionality," in Proceedings of SICE-ICASE International Joint Conference (2006) pp. 1081–1084.
[26]H. Lee, D. Kang, and W. Moon, "Design and fabrication of the high directional ultrasonic ranging sensor to enhance the spatial resolution," in Proceedings of Transducers and Eurosensors '07 (2007) pp. 1333–1336.
[27]S. Kim and Y. Kim, "Robot localization using ultrasonic sensors," in the Proceedings of 2004 IEEE/ RSI International Conference on Intelligent Robots and Systems (2004), Vol. **4**, pp. 3762–3766.
[28]Sv. Noykov and Ch. Roumenina, "Calibration and interface of a Polaroid ultrasonic sensor for mobile robots," Sens. Actuators, A **135**, 169–178 (2007).
[29]M. B. Moffett and R. H. Mellen, "Model for parametric acoustic sources," J. Acoust. Soc. Am. **61**, 325–337 (1977).
[30]B. K. Novikov, O. V. Rudenko, and V. I. Timoshenko, *Nonlinear Underwater Acoustics* (Acoustical Society of America, Melville, 1987), pp. 48–97 and 135–197.
[31]T. F. Hunter and R. H. Bolt, *Sonics* (Acoustical Society of America, Melville, 2000), pp. 86–126.
[32]P. C. Pedersen, O. Tretiak, and P. He, "Impedance-matching properties of an inhomogeneous matching layer with continuously changing acoustic impedance," J. Acoust. Soc. Am. **72**, 327–336 (1982).
[33]W. P. Mason, *Electromechanical Transducers and Wave Filters* (Van Nostrand Reinhold, New York, 1942), pp. 158–184.
[34]C. H. Sherman and J. L. Butler, *Transducers and Arrays for Underwater Sound* (Springer, New York, 2007), pp. 384–386 and 551–552.
[35]R. L. Pritchard, "Mutual acoustic impedance between radiators in an infinite rigid plane," J. Acoust. Soc. Am. **32**, 733–737 (1960).
[36]G. S. Garrett, J. N. Tjøtta, and S. Tjøtta, "Near field of a large acoustic transducer. Part II: Parametric radiation," J. Acoust. Soc. Am. **74**, 1013–1020 (1983).
[37]N. T. Adelman and Y. Stavsky, "Flexural-extensional behavior of composite piezoelectric circular plates," J. Acoust. Soc. Am. **67**, 819–822 (1980).
[38]S. H. Chang and B. C. Du, "Optimization of asymmetric bimorphic disk

transducers," J. Acoust. Soc. Am. **109**, 194–202 (2001).
[39]B. Belgacem, F. Calame, and P. Muralt, "Thick PZT sol-gel films for pMUT transducers performances improvement," in Proceedings of the 2006 IEEE Ultrasonics Symposium (2006), pp. 922–925.
[40]J. Yang, S. Kim, J. Yeom, C. Koo, C. Hwang, E. Yoon, D. Kim, and J. Ha, "Piezoelectric and pyroelectric properties of $Pb(Zr,Ti)O_3$ films for microsensors and actuators," Integr. Ferroelectr. **54**, 515–525 (2003).
[41]H. Lee, J. Tak, W. Moon, and G. Lim, "Effects of mutual impedance on the radiation characteristics of transducer arrays," J. Acoust. Soc. Am. **115**, 666–679 (2004).
[42]K. Misu, T. Nagatsuka, S. Wadaka, C. Maeda, and A. Yamada, "Film bulk acoustic wave filters using lead titanate on silicon substrate," in Proceedings of the 1998 IEEE Ultrasonics Symposium (1998), pp. 1091–1094.
[43]S. S. Rao, *The Mechanical Vibrations* (Addison-Wesley, Reading, MA, 1995), pp. 152–155 and 200–205.
[44]M. Greenspan, "Piston radiator: Some extensions of the theory," J. Acoust. Soc. Am. **65**, 608–621 (1979).
[45]Technical documentation from Brüel & Kjær, "Microphone handbook," http://www.bksv.com (Last viewed 11 March 2008).
[46]K. Ishihara, Y. Chi-Fan, A. A. Ayon, and M. A. Schmidt, "An inertial sensor technology using DRIE and wafer bonding with interconnecting capability," J. Microelectromech. Syst. **8**, 403–408 (1999).
[47]H. Ju and Y. Kim, "Directional characteristics of parametric loudspeakers in near-field" (in Korean), in Proceedings of the KSNVE 2005 Annual Autumn Conference (2005), pp. 545–550, available from Korea Institute of Science and Technology Information, http://www.kisti.re.kr/KISTI/index.jsp (Last viewed 11 March 2008).
[48]M. F. Hamilton and D. T. Blackstock, *Nonlinear Acoustics* (Academic, San Diego, 1998), pp. 76–84.
[49]G. S. Garrett, J. N. Tjøtta, and S. Tjøtta, "Near field of a large acoustic transducer. Part III: General results," J. Acoust. Soc. Am. **75**, 769–799 (1984).
[50]D. R. França and A. Blouin, "All-optical measurement of in-plane and out-of-plane Young's modulus and Poisson's ratio in silicon wafers by means of vibration modes," Meas. Sci. Technol. **15**, 859–868 (2004).

On the Larsen effect to monitor small fast changes in materials

Oleg I. Lobkis and Richard L. Weaver[a)]
Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green, Urbana, Illinois 61801

(Received 22 May 2008; revised 16 January 2009; accepted 22 January 2009)

At sufficient gain an ultrasonic feedback circuit rings with a "Larsen" tone that depends on the acoustic properties of the solid body to which it is attached. Because the frequency of this tone may be measured virtually continuously and with high precision, it is potentially capable of responding to fast small changes in materials. Here a tentative theory for Larsen dynamics is introduced and compared with laboratory measurements. Larsen monitoring is then applied to observation of the curing process of a cement paste sample and to studies of "slow dynamics" in which mesoscale nonlinear materials subjected to modest loads experience a drop in modulus but then recover in a characteristic manner like $\log(t)$. The present technique, using as it does higher frequencies and the Larsen effect and brief transient loads, extends investigations of slow dynamics to earlier times. For the materials and loads investigated, it is found that $\log(t)$ behavior fails at the shortest times, recovery being faster over the first several milliseconds. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3081530]

PACS number(s): 43.25.Ts, 43.35.Yb, 43.38.Ew, 43.40.Le [TDM] Pages: 1894–1905

I. INTRODUCTION

It has recently been suggested that measurement of the frequency of direct ultrasonic feedback in an irregular reverberant body could permit monitoring of materials or structures. The Larsen effect[1] in which such feedback results in a single tone is familiar to all. Weaver and Lobkis[2] showed that the width of the steady tone arising from such a nonlinear system can be extremely small and that the frequency of that tone is sensitive to tiny changes in a structure. Detection of changes of a part in 10^8 was shown to be not difficult, and detection of smaller changes is possible in principle. As change in electronics and temperature can also contribute, the technique is best applied to the detection of sudden changes that such drifts are unlikely to generate.

The usual audiofrequency experience is of a single tone, which appears when gain is above a critical level. The tone waxes as amplifier gain is adjusted above that threshold and wanes or disappears when the gain is decreased. It was argued by Weaver and Lobkis[2] that the intrinsic width of the tone was due to background noise and was, in their measurements, too small to measure unless that noise was artificially augmented. It was also shown that the tone was highly stable, varying only due to imperfectly controlled environmental conditions such as temperature. The same considerations apply to the stability of the ultrasonic oscillators used for clocks, and so one imagines that achieving stability greater than that demonstrated in Ref. 2 is technically straightforward.

The frequency of such a tone is easy to measure with high precision and is sensitive to tiny changes in materials. This suggests new methods for nondestructive evaluation structural health monitoring and materials characterization. Here we discuss theory for unstable linear feedback in reverberant acoustic bodies and theory for the resulting limit cycle. We introduce an application to characterization of a material cure and to the phenomenon known as slow dynamics.

The past decade has seen several reports[3–10] of a provocative behavior in materials with mesoscale nonlinearity. These include sandstones and limestones and cements that have complex microstructures composed of imperfectly sintered grains. The nonlinearities observed in these materials are stronger than in classically nonlinear materials. They are generally not quadratic, and they show a memory effect in which loads applied and then removed at one time affect tangent moduli at later times. A typical test from this group will apply a continuous harmonic load to a sample at frequencies of a few kilohertz and will do so with modest strains of the order of 10^{-6}. The load is then released, and the stiffness of the sample is measured and found to be diminished relative to its original stiffness by a fraction proportional to the applied strain. The modulus is then observed to recover slowly, over minutes, and to do so in a manner proportional to the logarithm of time since the load was removed. This is technically a nonlinearity, as modulus depends on load, but it is a nonlinearity of non-classical form, with a slow history dependence. The $\log t$ behavior appears to be universal; it is seen in a variety of materials. It is not understood.

This literature has observed the effect in many materials and at various loads. When temperature is well controlled, the stiffness at late times is identical to the stiffness before the load was applied; the material recovers completely. That it recovers is perhaps unsurprising, as the strains were modest. That its recovery is proportional to $\log t$ over a substantial range in time from 1 to 1000 s or more is intriguing.

The slow-dynamics literature has developed various clever methods for high-precision measurements of the ma-

[a)]Author to whom correspondence should be addressed. Electronic mail: r-weaver@uiuc.edu

 0001-4966/2009/125(4)/1894/12/$25.00

terial changes. One way is to monitor the change in resonant frequency. Another is to monitor response to a weak harmonic excitation at a frequency near a resonance. If that frequency is on the shoulder of a vibration resonance, then a change in material properties manifests as a change in amplitude. Both these methods enhance measurement sensitivity by factors of order Q, the quality factor of the resonance. These techniques have so far required a second or more to complete, thus precluding study of slow dynamics immediately after the load is removed. The methods are furthermore poor at short times in that the continuous loading makes problematic the identification of the zero of time needed for an unambiguous identification of log t. For these reasons we conjecture and demonstrate in Sec. IV that Larsen monitoring may complement these methods.

Larsen monitoring is similar to more complex feedback techniques that have existed in the acoustics literature for some time.[11] Ultrasonic sing-around, in which a specified zero crossing in a received transient signal is used to trigger the next transient source, can be used to measure changes in propagation time with exceptional precision. Phase-locked resonance tracking uses a voltage controlled oscillator and a lock-in amplifier, essentially a phase-locked loop, to measure natural frequencies of samples with exceptional precision. In both cases the frequency that the analog nonlinear feedback system chooses is interpretable in terms of a physical quantity, a propagation time, or a natural frequency.

A Larsen circuit uses a simpler, nominally linear, direct feedback for which the source is an algebraic function of the instantaneous input. Because the Larsen field is reverberant, with wavelengths short compared to system size and absorption times long compared to transit times and with multiple complex acoustic paths between source and receiver, the frequency that the Larsen circuit chooses is not readily interpreted as a physical quantity of interest. Only its changes are noteworthy. These changes correspond to changes in material properties at any or all locations within the sample to which the reverberant field penetrates.

In Sec. II a theory for a single Larsen line is formulated. This includes predictions for the linear stability of a quiescent circuit, the resulting steady state screech and its fluctuations, and its sensitivity to changes in material or circuit parameters. Laboratory experiments, both supportive of the theory and not so, are presented in Sec. III. Then in Sec. IV we apply Larsen monitoring to the characterization of a cure evolution and to the slow dynamics of a material sample with mesoscale nonlinear elasticity as it recovers from transient damage,

II. LARSEN CIRCUIT: THEORY

Consider the electro-mechanical circuit illustrated in Fig. 1 in which a waveform $A(t)$ enters a source transducer ("actuator") and is propagated through the sample by an electro-mechanical Greens' function G to a receiver, producing a signal $S(t)$, which is then filtered, amplified [and potentially augmented by amplifier noise $n(t)$], and fed to the actuator. A voltage divider is illustrated that permits control

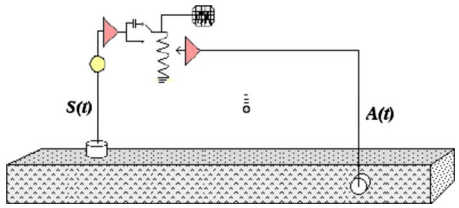


FIG. 1. (Color online) Sketch of test configurations. An electronic feedback circuit filters the signal S with high-pass and band-pass filters (we use passive four-pole filters from TTE) and amplifies it. On occasion it passes then through a switch that inserts extra capacitance. The signal is then passed through an adjustable voltage divider and amplified further before being passed to a piezoelectric actuator at A that provides the ultrasonic source. In some cases the second amplifier is replaced with a comparator, and the system maintained in a regime in which the first amplifier is strictly linear. The circuit typically used two Panametrics model 56000 preamplifiers, each switchable between 40 and 60 dB. Occasionally a third transducer (not shown) monitors the acoustic state of the body. The sample pictured is composed of cured cement paste and has dimensions of $2.5\times2.5\times28$ cm^3. Transducers, typically Physical Acoustics's WD 492 and SNAA04 Acoustic emission sensors, with diameters of 15 and 17 mm and useful frequency ranges from 200 to 700 kHz corresponding to wavelengths of several mm, were placed on different sides of the sample so that their direct fields did not overlap. Typical amplitudes at A are a few volts. Typical signal strengths at S are millivolts or less. Signals were captured and digitized by a LeCroy digital oscilloscope for visual monitoring and by a National Instruments PC-based digitizer with deep (2^{19} words) memory for analysis. Circuit design details are generally unimportant. As described following Eq. (1), a Larsen signal is observed if feedback gain g is sufficient to overcome the electro-mechanical insertion loss between the transducers.

of gain. This simple circuit differs from complex self-oscillating feedback circuits used elsewhere in acoustics[11] and which in principle could be used here as well; it is not *a priori* clear what the relative advantages and disadvantages would be. In the present application, the linear acoustic response G is dominated by a diffuse reverberant field rather than direct propagation. Its Fourier transform, defined by $G(\omega)=\int G(t)\exp\{-i\omega t\}dt$, is an irregular random-looking function of ω with correlations over the range of the inverse of an absorption time. The envelope of $G(\omega)$ is defined by the filters, transducers, and amplifiers. As described below, G's uncontrolled irregular details determine the self-oscillation frequency of the circuit; they are unrelated to any gross material properties.

A. Linear regime

Governing equations for the linear regime, i.e., at signal amplitudes below amplifier saturation, are

$$A(t) = g \otimes S(t),$$

$$S = G \otimes A = G \otimes g \otimes S. \tag{1}$$

The symbol $\otimes$ represents a temporal convolution. Equation (1) predicts unstable growth $S\sim\exp(i\gamma t)$ at (complex) frequency $\gamma=\omega+i\nu$ such that $G(\omega+i\nu)g(\omega+i\nu)=1$. This regime was discussed in Ref. 2, with special attention to the effects of noise. The conclusion suggests, but not compellingly, that the frequency of the ultimate steady state screech is the real part of γ, at least if gain is just sufficient for linear instability such that ν is small and negative.

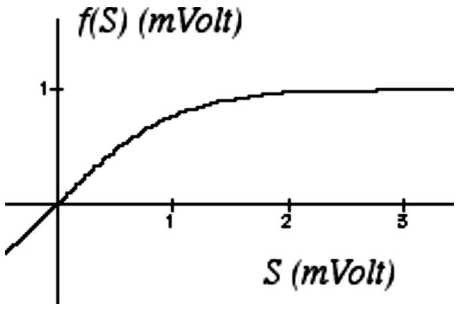


FIG. 2. The nonlinear behavior of the amplifiers is presumed to be smooth, with saturation at an input of several millivolts.

B. Nonlinear circuit

In the nonlinear regime, with saturation, we replace the first of the above equations with

$$A(t) = g \otimes [f(S)], \tag{2}$$

where f is a (odd in S is plausible) function of the signal level S. We also assume that f is instantaneous; there are no nonlinear time delays. Full saturation occurs where $f'(S)=0$. To ensure consistency with the above presumed form in the linear regime, we take $f'(0)=1$. Typically we might expect f to take a form like that illustrated in Fig. 2 with unit slope at small input signal and saturation for input signals greater than a few millivolts. A good choice of f might be

$$f = (F_o/|g|)\tanh(|g|S/F_o), \tag{3}$$

with amplitude of $F_o \sim 5$ V, a typical maximum output of our preamplifiers, and $|g|$, an experimentally adjustable parameter that varies with gain, between 100 (40 db) and 1 000 000 (120 dB). G varies with frequency and sample volume and attenuation. It is independent of gain. Here it is typically of order 10^{-3}, so that about 60 dB is required for the fixed point at $S=0$ to go unstable. The large g limit is simple; for this case $f=(F_o/|g|)\mathrm{sgn}(S)$. This corresponds to the use of a *comparator*[12] rather than a conventional amplifier.

Then, with H defined by

$$H \otimes G \otimes g = \delta(t), \quad \tilde{H}(\omega) = 1/\tilde{G}(\omega)\tilde{g}(\omega), \tag{4}$$

we find that the circuit is governed by the nonlinear time-domain convolution equation

$$H \otimes S = f(S(t)). \tag{5}$$

C. Steady state screech

We seek steady solutions to the above and are particularly interested in the frequencies of such steady solutions. We further seek to understand how sensitive such a solution is to perturbations in material properties, as this is the basis for the imagined applications. It is also desirable to understand the stability of the steady solutions, i.e., the circuit's behavior when it is suddenly perturbed; how quickly does it relax back to the steady solution?

We take S in the form of an oscillation with slowly varying complex amplitude $V(t)$,

$$S(t) = V(t)\exp(i\Omega t) + \text{c.c.} = 2|V(t)|\cos(\Omega t - \phi(t)), \tag{6}$$

where the slowly varying $|V(t)|$ is (half) the instantaneous amplitude of the oscillation, ϕ is $\arg(V)$, and Ω is the nominal frequency of $S(t)$.

If $V(t)$ varies slowly compared to the time delays in H (often equal to acoustic dwell times in the body, i.e., to absorption times), it is helpful to write Eq. (5) as (see the Appendix for a derivation)

$$\exp(i\Omega t)[\tilde{H}(\Omega - id/dt)]V(t) = V\exp(i\Omega t)q(|V|) + hft's. \tag{7}$$

We shall neglect the high frequencies ($hft's$) as multiples of Ω and therefore outside the pass band of Gg. $q(|V(t)|)$ is a function formed from f by extracting the term that is linear in V. q is given in terms of a Fourier integral:

$$q(V_o) = \frac{1}{2\pi V_o}\int_0^{2\pi} \exp(-i\tau)f(2V_o\cos\tau)d\tau$$

$$= \frac{i}{2\pi V_o}\int_0^{2\pi} \exp(-i\tau)f(2V_o\sin\tau)d\tau. \tag{8}$$

If $f(S)$ is an odd function, then q is real. For the special case $f=F_o\,\mathrm{sgn}(S)/|g|$, we find $q(V)=2F_o/\pi|g|V$.

Equation (7) has a steady solution $V(t)=V_o$ given implicitly by

$$q(|V_o|) = \tilde{H}(\Omega), \tag{9}$$

thus implying that the frequency Ω of auto-oscillation must be such that $H(\Omega)$ is real and positive. This is a familiar condition, that the phase gain on any one full pass through the circuit is a multiple of 2π. The conclusion $\mathrm{Im}\,H(\Omega)=0$ is testable in the laboratory; it is confirmed below.

D. Time-dependent perturbations

Now consider a time-dependent V near the above limit cycle, with $V=V_o+u(t)$ and small complex $u(t)$. Equation (7) becomes

$$[\tilde{H}(\Omega - id/dt)][V_o + u(t)] = q(|V_o + u(t)|)[V_o + u(t)]. \tag{10}$$

At leading order in u, this is

$$[\tilde{H}(\Omega - id/dt)]u(t) = q(V_o)u + V_o q'(V_o)\mathrm{Re}\,u. \tag{11}$$

We seek solutions of the form $u=|U|\exp(i\alpha+\lambda t)$. Thus

$$[\tilde{H}(\Omega - i\lambda)] = q(V_o) + V_o q'(V_o)\mathrm{Re}(e^{i\alpha+\lambda t})e^{-i\alpha-\lambda t}. \tag{12}$$

The time dependence drops out only if λ is real. We then deduce

$$[\tilde{H}(\Omega - i\lambda) - \tilde{H}(\Omega)] = V_o q'(V_o)e^{-i\alpha}\cos\alpha, \tag{13}$$

which may be considered an equation for the real quantities λ and α. The equation may be rewritten by eliminating α:

$$|[\tilde{H}(\Omega - i\lambda) - \tilde{H}(\Omega)]|^2$$

$$- V_o q'(V_o)\mathrm{Re}[\tilde{H}(\Omega - i\lambda) - \tilde{H}(\Omega)] = 0. \quad (14)$$

This is an implicit equation for the (real) λ. If λ is small compared to the scale over which H varies (i.e., $\lambda \ll 1/$absorption time), then the difference $H(\Omega - i\lambda) - H(\Omega)$ may be written as $-i\lambda H'(\Omega)$ and an explicit expression derived for λ

$$\lambda = V_o q' \ \mathrm{Im}\ H'/|H'|^2 = (V_o q'/q) H \ \mathrm{Im}\ H'/|H'|^2$$

$$= (V_o q'/q) XY'/(X^2 + Y^2), \quad (15)$$

where $H = 1/Gg = (X + iY)$. The first factor in parentheses in the last term of Eq. (15) is real and dimensionless and of order −1, so we conclude that the Larsen state is stable ($\lambda < 0$) if and only if $\mathrm{Im}\ H' > 0$. We also conclude that

$$\alpha = \arg H'(\Omega) \quad (16)$$

plus any integer multiple of π.

Equation (16), together with the conditions $\mathrm{Im}\ H' > 0$, implies that this perturbation of the Larsen state has an amplitude $V_o + \mathrm{Re}\ u = V_o + |U|\cos\alpha \exp(\lambda t)$. Similarly the perturbed state has a frequency $\Omega + \lambda(|U|/V_o)\sin\alpha \exp(\lambda t)$. The ratio of frequency perturbation to fractional amplitude perturbation is $\lambda \tan\alpha$.

The value of λ predicted by Eq. (15) is of the order of scale of the frequency variations in H. So we expect the assumption $\lambda \ll 1/\tau_{\mathrm{absortption}}$ made following Eq. (14) to be invalid and corresponding predictions to be non-rigorous. Nevertheless, the above gives estimates for λ and α. More precise values must be found by numerical analysis of Eq. (14) based on analytic continuations of H.

E. Time-independent perturbations

We wish also to estimate the effect of a change in material properties on Larsen frequency. If the material perturbation is large, it may well be that the system will jump to an entirely different Larsen state about which little can be said. If the material perturbation is small enough, we expect a comparably small change in Ω for which theoretical insight may be possible. It is helpful to distinguish two parameter regimes, of high and low modal overlaps.

At low overlap, where G is a superposition of contributions from distinct normal modes, the Larsen state will be dominated by a single normal mode r of the sample. In this regime, we may write gG in the form

$$gG \sim B(\omega)/[(\omega - \omega_r) + i\varepsilon_r],$$

where the factor B has a magnitude and phase inherited from the amplifiers and filters, $B = |B|\exp\{i\phi\}$. These vary only slowly with ω. The mode has natural frequency ω_r and a quality factor $Q = \omega_r/2\varepsilon_r$. The Larsen frequency (the solution of $\mathrm{Im}\ gG = 0$) is then

$$\Omega = \varepsilon_r/\tan\phi + \omega_r.$$

Small variations in elastic or inertial properties generate comparable changes in ω_r and relatively negligible changes in Q. Thus the Larsen frequency will track changes in the sample's natural frequencies. Appealing to standard first-order time-independent perturbation theory for normal modes,[13] it may be concluded that increased mass leads to negative semi-definite changes in Larsen frequency. Similarly, decreases in stiffness will lead to negative semi-definite changes in Larsen frequency.

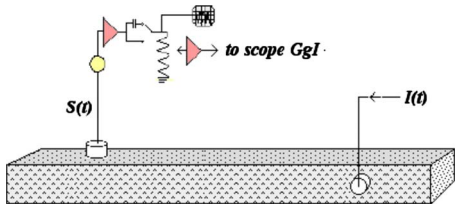


FIG. 3. (Color online) The passive linear response Gg corresponding to Fig. 1 is measured by applying an impulse $I(t)$ to the actuator and measuring the response GgI of the linear circuit.

At high overlap but still in a reverberant regime, gG is the Fourier transform of a causal decaying (but otherwise stationary) Gaussian random process. The resulting statistics on $|gG(\omega)|^2$ are termed Ericson fluctuations.[14] gG is a complex Gaussian random process with correlation range equal to the inverse of an absorption time. Its phase generally increases with frequency, but not strictly so. How its phase depends on changes in material properties is not at all clear. Addition of mass surely must, in some average sense, delay the responses (it lowers wave speeds) and thus augment phase and diminish Larsen frequency. But that is an argument that pertains to an average. We have been unable to reliably conclude that Larsen frequency has the same kind of monotonicity at high overlap that it ought to have at low overlap.

III. MEASUREMENTS OF THE LARSEN SIGNAL: FUNDAMENTALS

Figure 1 illustrates the experimental configuration. A piezoelectric sensor is placed at an arbitrary position $\boldsymbol{S}$ on an irregular elastic body. The signal from the sensor at S is filtered to control the frequency range available to the feedback circuit, amplified, sampled by a digitizer, passed through an adjustable voltage divider, and amplified further before being passed to a piezoelectric actuator at A that provides the ultrasonic source. Sometimes the filter is placed *after* the amplifiers. For reasons described below, sometimes the second amplifier is replaced with a comparator. Turning on the amplifiers leads to a time-domain signal in which a nominally single-frequency signal grows until the amplifiers saturate.

A switch is sometimes inserted, as illustrated, in order to impose a controlled sudden electronic perturbation by adding or deleting a small capacitance. Figure 1 also illustrates the dropping of a small ball to induce a perturbation to the mechanical system in order to investigate slow dynamics (Sec. IV). Figure 3 illustrates the measurement of the linear part of the transfer function, $Gg = 1/H$, needed in Eqs. (9)–(16).

A. Steady state screech

Section II presented some testable theoretical predictions and speculations. In particular it predicted that the

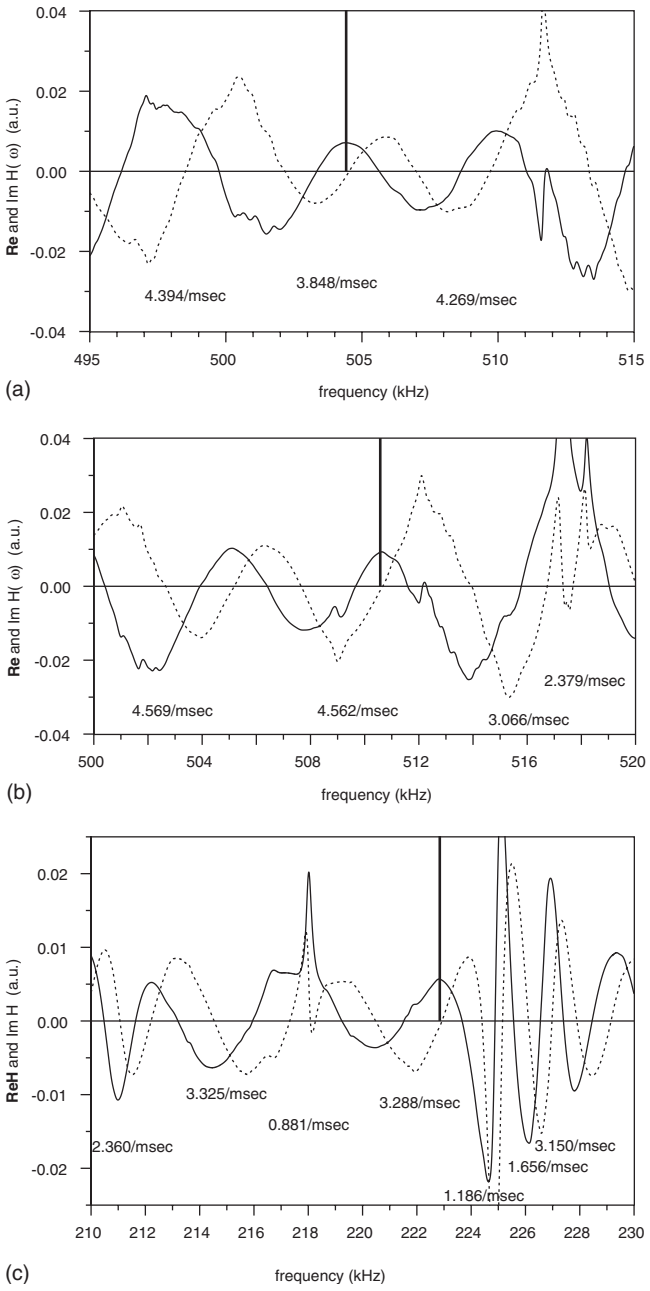


FIG. 4. (a) A span of the spectrum near 500 kHz in a cement paste sample. The bold vertical line indicates the position of the Larsen line. (b) The spectrum H and Larsen frequency of a different configuration. (c) A low-pass filter brings the Larsen state to a frequency of 222.8 kHz.

Larsen frequency Ω is such that Im $H(\Omega)=0$ and Re $H>0$. It also predicted that the Larsen signal, when perturbed, relaxes to its steady state like $\exp(-|\lambda|t)$ with a specified ratio of perturbed frequency and amplitude. Both λ and α are given theoretically in terms of the Greens' function in the vicinity of Ω.

Figure 4 gives some sample spectra $H(\omega)$ as obtained conventionally in the configuration illustrated in Fig. 3 by applying an input pulse $I(t)$ at A and measuring the response $v(t)$ equal to the convolution of Gg and I in a low amplitude regime in which there is no saturation. H was defined as the ratio of their Fourier transforms. $\tilde{H}(\omega)=\tilde{I}(\omega)/\tilde{v}(\omega)$. Superposed on these spectra are vertical lines indicating the frequency of the Larsen screech.

Three cases are shown, corresponding to different transducer positions or different filters. In all cases Im H (the dashed line) vanishes at the Larsen frequency, in accord with Eq. (9). Similarly Re H (solid line) is in all cases positive. Figure 4(a) shows a span of the spectrum near 500 kHz in a cement paste sample. The bold vertical line indicates the position of the Larsen line. Apart from a slight frequency shift that is attributable to temperature drift and/or trigger jitter, the Larsen state has a frequency at which Im $H=0$ and Re $H>0$, as required by theory. There are three other points in this span for which Im $H=0$ and Re $H>0$ also. Each is labeled by its value of $-\lambda$ according to Eq. (15). All are positive, indicating that each is stable (i.e., $\lambda<0$). The labels also show that this Larsen state occurs at the least stable of these points, although the differences are slight. One conjectures that the smaller value for H in the vicinity of 504 kHz, corresponding to a larger value of Gg, may generate stronger linear instability of the quiescent state and may thus be responsible for the system choosing this state rather than one of the others. Alternatively, it may be that the simple theory [Eqs. (6)–(16)] is inadequate for a good estimate of λ. A value of λ of order 4/ms corresponds to an (imaginary) frequency difference of 0.65 kHz, a range over which H is imperfectly approximated as linear.

Figure 4(b) shows the spectrum and the Larsen frequency for different circuit impedances and transducer positions but the same sample and filters. The Larsen state is similar to that of Fig. 4(a). This time the Larsen state chooses the most stable of the candidate frequencies and also the one with the greatest Gg.

Figure 4(c) shows the spectrum and Larsen frequency after the high-pass filter is removed, thus bringing the Larsen state to a frequency of 222.8 kHz. Again one notes a slight shift between the Larsen frequency and the position of the corresponding zero in Im H, again ascribable to temperature drift and/or trigger jitter. In this range of frequencies, there are seven candidate points at which the criteria Im $H=0$ and Re $H>0$ are met. The Larsen dynamics has chosen that frequency, which is among the most stable, and has the greatest value of $gG=1/H$.

B. Time-dependent perturbations

On switching in or out a minor perturbation to the electronics, as illustrated in Fig. 1, we can measure circuit dynamics in the vicinity of the limit cycle. The switching is effected electronically using transistors and takes place in microseconds. Figure 5(a) shows the rms Larsen amplitude in the cement paste sample after this is done. The Larsen signal is seen to evolve after the perturbation with a time scale that is of the order of the absorption, here about 200 μs.

An estimated instantaneous frequency of the signal whose amplitude is given in Fig. 5(a) was determined by a running average of its zero spacings and is shown in Fig. 5(b). The relaxation to the new frequency is, as is the amplitude in Fig. 5(a), accomplished on the scale of the absorption time in the sample. These dynamics are believed to be linear [as assumed in the derivations of Eqs. (10)–(16)] because a

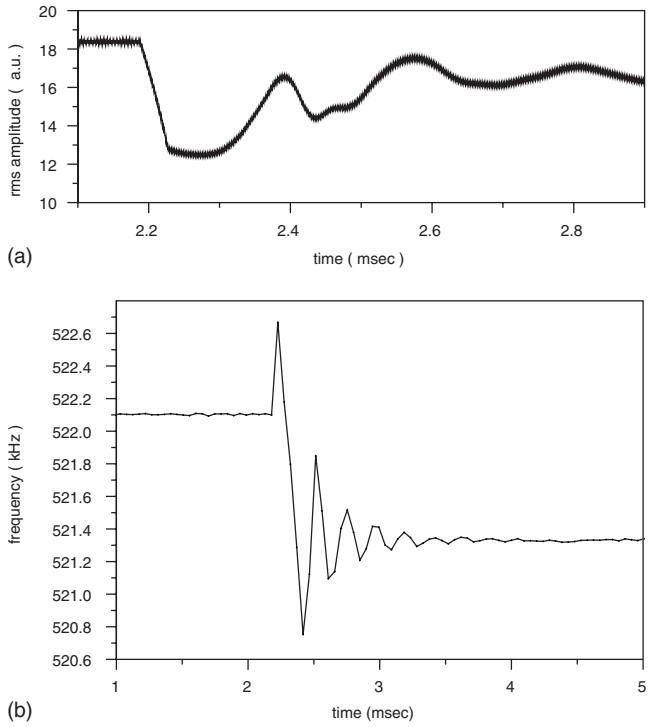


FIG. 5. (a) The Larsen signal's root mean square amplitude (smoothed over a time period of 0.5 ms) relaxes back to a new steady state after a change in impedance is applied at time of 2.2 ms. (b) The instantaneous frequency of the Larsen signal of (a).

repeat test with a greater impedance change merely rescales figures like Fig. 5(b) without affecting their shapes. One can also evaluate the phase accumulation during the transition, $\int[\omega(t)-\omega(\infty)]dt$. It is less than unity, so again linearity is plausible. The decaying oscillations are intriguing, especially as that behavior is not consistent with the theory [Eqs. (10)–(16)], which predicted real λ. The theory appears to be incorrect or inapplicable. The time scale for the relaxation is, however, of the same order as that predicted by theory. One may speculate, inasmuch as the many solutions of Eq. (9) are separated by frequency differences comparable to the beat frequency above, that the oscillations in Fig. 5(b) are such beats.

The above behavior was obtained with the use of a comparator in place of the second amplifier. The use of a pair of conventional ultrasonic preamplifiers (Panasonic 5660B) sometimes allowed perturbations that relaxed hundreds of times more slowly. This is not understood, but these preamplifiers are not well characterized in their saturation regime, nor are they guaranteed to have no nonlinear delay. The advantage of a comparator, which puts out a signal $f(S)$ essentially equal to $F_o\,\mathrm{sgn}(S)$, is that the nonlinearity and saturation behavior are understood and well controlled. Our comparator puts out a signal F_o with a fixed amplitude of about 5 V; it switches at a threshold of ±20 mV, with a switching time of several nanoseconds.

Signal dynamics were also studied in an aluminum bar of dimensions $35\times2.5\times2.5$ cm^3. Absorption in the aluminum was about three times slower than in cement (and dominated by losses into the transducers, not intrinsic mechanisms.) Figure 6 shows the transfer function for this system

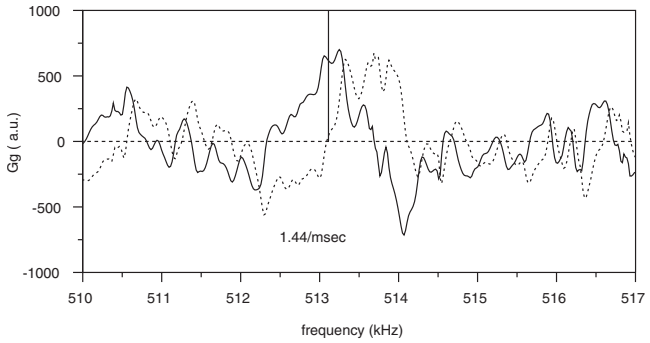


FIG. 6. A Green's function for an aluminum bar in the vicinity of its Larsen frequency.

in the vicinity of the Larsen line at 513.112 kHz. Again the Larsen state has chosen a frequency at which Im Gg vanishes and Re Gg is positive. The line has been labeled with the theoretical stability parameter $\lambda=1.44/\mathrm{ms}$ calculated for this point. Figure 7 shows the instantaneous frequency after a perturbation corresponding to several hertz. Phase perturbation $\int[\omega(t)-\omega(\infty)]dt$ is still small, and acoustic parameters such as modal density and absorption time are similar, yet the behavior is very different. We see that the perturbed signal relaxes to the new equilibrium in an irregular (yet reproducible) manner, but with a time constant of 10 ms or more, much greater than the predicted $\lambda\sim0.69$ ms. The signal dynamics is qualitatively different. Neither system has a behavior that comports with theory. The enormous phase space available in such structures appears to lead to some nonobvious behaviors that, so far, resist theoretical understanding.

C. Time-independent perturbations

The sign of the change in Larsen frequency upon adding mass was also investigated. As in Ref. 2 drops of water, from 1 to 10 mg, were added to the surface of aluminum and cement paste samples (the latter to an applied hydrophobic surface). All tests were conducted in the high modal overlap regime in which a large number of natural modes contribute at any given frequency. These tests were done on a variety of Larsen states at various different frequencies and different transducer positions. In all cases it was found that the origi-

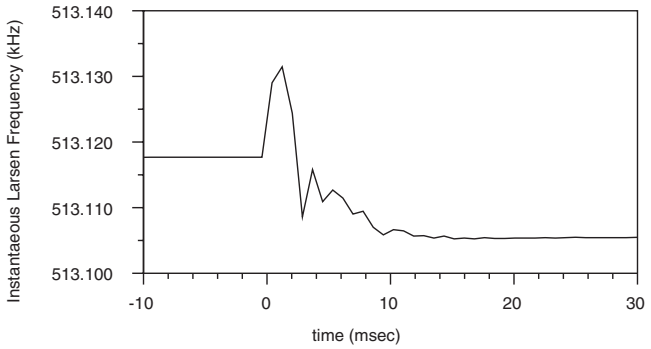


FIG. 7. Larsen dynamics in the aluminum bar. Instantaneous frequency was determined by a chi-square fit to time-windows of length of 0.8192 ms. Amplitude shows similar although somewhat less erratic behavior, varying by about 10%.

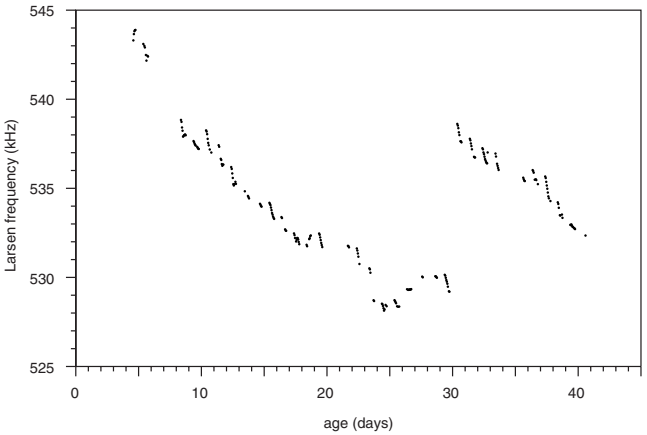


FIG. 8. The Larsen frequency of a sample of cement paste in the weeks following its creation shows a steady slight decrease at about 3% per month. Larsen frequency was assessed several times on each of many days. Within each day we observe a downward drift as sunlight heated the room. The large jump at day 30 is due to a loss of Larsen limit cycle stability and a jump to a new limit cycle; it is not indicative of a correspondingly large (2%!) change in the structure.

nal Larsen state was recovered when the water drop was removed. In no case was the sign of the change in Larsen frequency consistent. Increases were as common as decreases. This is in accord with the theoretical expectations (Sec. II C) that there should be no such monotonicity unless the modes are distinct. The magnitudes of the fractional changes were, as might have been anticipated, comparable to the ratio of added mass to the sample mass.

The change in Larsen signal with ambient temperature or degree-of-material cure was also studied. As seen previously, increased temperature, corresponding as it does in these materials to nearly uniformly decreased wave speed, induced greater delay and lower frequency. Thus Larsen frequency was typically found to evolve slowly and continuously as ambient temperature evolved, by parts in 10^5/min corresponding to small fractions of a degree per minute.

IV. APPLICATIONS

As an illustration of the potential utility of Larsen monitoring, we have used it to study the cure of a cement paste sample and to study the slow dynamics discussed in Sec. I.

A. Cure monitoring

Larsen frequency was monitored in a sample of cement paste in the weeks following its manufacture. At each of several times in the course of each weekday for a month, Larsen frequency was measured at 2 s intervals for 100 s at a time. As seen in Fig. 8 Larsen frequency showed a steady decrease from day to day. This change is attributable to curing; it tracks changes in resonance frequency near 5000 Hz of a sister sample. The steady decrease is modulated by a daily fluctuation attributable to variation in laboratory temperature. At day 30 it jumped autonomously by a large amount [comparable to the spacing between zeros of the imaginary part of the Green's function (see Fig. 4)]. This is presumably due to the Larsen state that it had enjoyed for 4 weeks, losing its stability (due to slow changes in material properties reaching a critical point) and the sample promptly re-settling in a new Larsen state.

Figures 9 show closeups of the Larsen frequency on 2 s intervals in one of the tests on day 15 and one of the tests on day 17. By chance, these exhibit sudden jumps. It is tempting to attribute the jump to a sudden change in material properties. The magnitude of the jump is tiny compared to the frequency scale of Gg; so the jump is unlikely to be like that seen between days 30 and 31. Interestingly, the jump of day 15 is followed by a slow recovery; the jump on day 17 is preceded by a similar trend. Few such changes were seen, but they are probably not rare, as the sample was monitored less than 1% of the time. It would be interesting to observe such changes while simultaneously listening for acoustic emissions.

B. Slow dynamics

Loading the sample by the transient impact of a wooden ball leads to a sudden shift in Larsen frequency that can be controlled. As with the autonomous shifts in Fig. 9, this shift then relaxes over a period of minutes. As described in the slow-dynamics literature,[3–10] application of moderate strains to materials with nonlinear mesoscale elasticity leads to an apparent damage in which the modulus diminishes but then recovers slowly, in a manner similar to that of Fig. 9. In an illustration of the potential usefulness of Larsen monitoring, we apply it here to the detection of such changes following the impact of small wooden balls dropped from a height of 20 cm on a cement paste sample. Concurrently with the ball

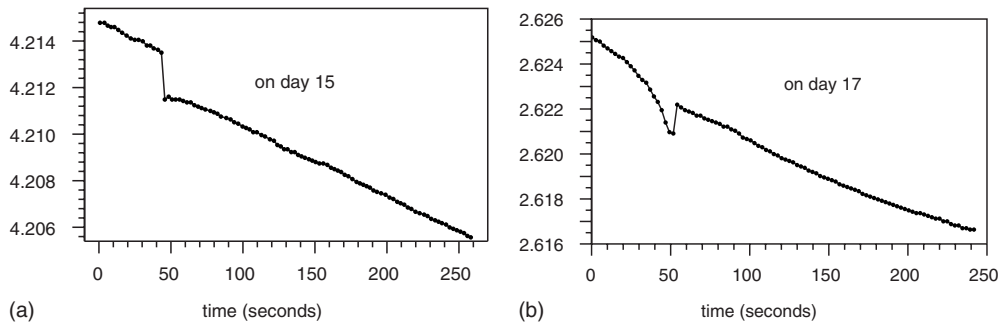


FIG. 9. On days 15 and 17 the cement paste sample exhibited sudden (within the 2 s resolution) changes in Larsen frequency (here plotted less 530 kHz) and slow recoveries. These changes are tiny compared to a shift associated with a change in Larsen state (e.g., Fig. 8 at day 30).

impact, we monitor a ~500 kHz Larsen frequency and find that it diminishes after the impact and then recovers slowly. Acoustic emission from the impact is largely confined to frequencies below 100 kHz and thus has little direct influence on the Larsen circuit, which incorporates a passive four-pole high-pass filter with a 3 dB point at 350 kHz. Our Larsen signals are typically captured at a sample rate of 2.5 MSa/s with 2^{19} 16-bit samples for a signal length of $T=210$ ms. To provide a baseline and a record of temperature drifts, Larsen signals are captured at regular time intervals for several minutes before the impact. The acoustic emission from the impact is then detected by a separate ultrasonic receiver and used as the zero of time and to trigger immediate Larsen monitoring for a period of $T=210$ ms after the impact. At later times the Larsen signals are again captured at non-coterminous finite intervals.

Each 2^{19} word sample is broken into 32 consecutive sub-samples and analyzed for its frequency. This is done in three different ways. One technique is to pad each sub-sample of length $T/32$ with a large number of zeros and then do a discrete Fourier transform. The result has a clear peak that can typically be resolved to a degree far finer than $32/T$. Alternatively, the record is fit by nonlinear least-squares to a sine wave of constant amplitude and phase. In some circumstances, a simple measurement of zero spacing suffices. The three methods were found to yield the same Larsen frequency.

Estimates of maximum loads and contact times may be made based on the familiar quasi-static Hertzian contact theory for elastic spheres.[15,16] We consider a ball of radius R, uniform mass density ρ, mass $m=(4\pi\rho/3)R^3$, Young's modulus E_1, and Poisson ratio ν_1. It impacts at speed $V_{\text{impact}}=\sqrt{2gH}$, where H is the height from which the ball is dropped. Johnson[15] tells us that the contact time T^*, the maximum penetration δ^*, the maximum pressure p^*, and the radius a^* of the contact region are given by (where $1/E=(1-\nu_1^2)/E_1+(1-\nu_2^2)/E_2$ and E_2 and ν_2 pertain to the half space)

$$\delta^* = (15\ \text{V}^2\ \text{m}/16\sqrt{RE})^{2/5}$$

(proportional to R and $H^{2/5}$, at constant ρ, E),

$$a^* = \sqrt{(\delta R)}$$

(proportional to R and $H^{1/5}$, at constant ρ, E)

$$p^* = 2E\delta^*/\pi a^*$$

(independent of R, and scaling with $H^{1/5}$ at constant ρ, E)

$$T^* = 2.87(\text{m}^2/RE^2\ \text{V})^{1/5}.$$

(proportional to R and to $H^{-1/10}$ at constant ρ, E).

Consider a representative case of a 5 mm radius wooden ball (with density of 0.8 g/cc) impacting on a cement half space at 2 m/s, for which $T^*=36\ \mu$s, $m=0.4$ g, $\delta^*=25\ \mu$m, $a^*=0.351$ mm, and $p^*=320$ MPa, for a maximum strain in the cement of 1.3%. The maximum stress is directly under the center of the contact and is compressive. For damage in materials such as cement and stone, the tensile stresses are probably the more important. It is the radial stress that is tensile; its maximum occurs at the edge of the contact zone and is $(1-2\nu)/3$ times the maximum compressive stress and thus considerably less than $|p^*|$. Radial stress then diminishes algebraically with further distance from the center of contact, like r^{-2}, and with depth on a scale of a. The volume that sees comparable tensile stresses thus scales like a^3, i.e., like R^3. It is plausible therefore to suppose that total damage scales with the cube of the radius of the ball.

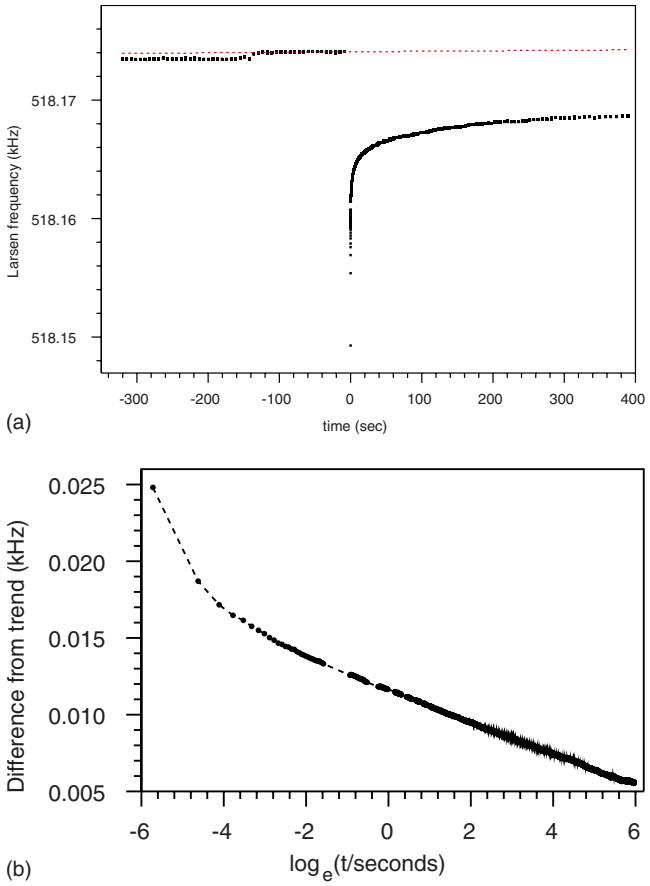


FIG. 10. (Color online) (a) The Larsen frequency of a cement sample is monitored for about 10 min, with continuous sampling at the most critical times. At time "0" a wooden ball impacts upon the sample. The Larsen frequency then drops by about 25 Hz and then recovers. The dotted line is a linear fit to the early time trend. The apparent jump at time −140 s is real, perhaps a small version of those seen in Fig. 9. (b) The difference between the Larsen frequency and the dotted line in (a) is plotted versus the log of the time after the impact.

Heat generated in the sample is bounded by the impact kinetic energy. The above wooden ball carries 0.0008 J for a specimen averaged temperature increase (we assume a 250 g specimen) of order of 10^{-6}°. The fractional modulus change associated with this temperature change is even smaller and negligible compared to the measured fractional frequency change. Thermal diffusivities in cements are of the order of 1 mm^2/s, so the temperature rise immediately under and after the ball impact is greater than this; at 1 s after impact it is about 1°.

A typical result is shown in Fig. 10. The Larsen frequency is seen to be stable for some minutes before impact, with a slow drift that can be ascribed to temperature changes. The observed variation of less than 1 Hz (out of 518.174 kHz) over 300 s corresponds to a temperature variation of 0.04°. The moment of impact is very clear in the plot, where

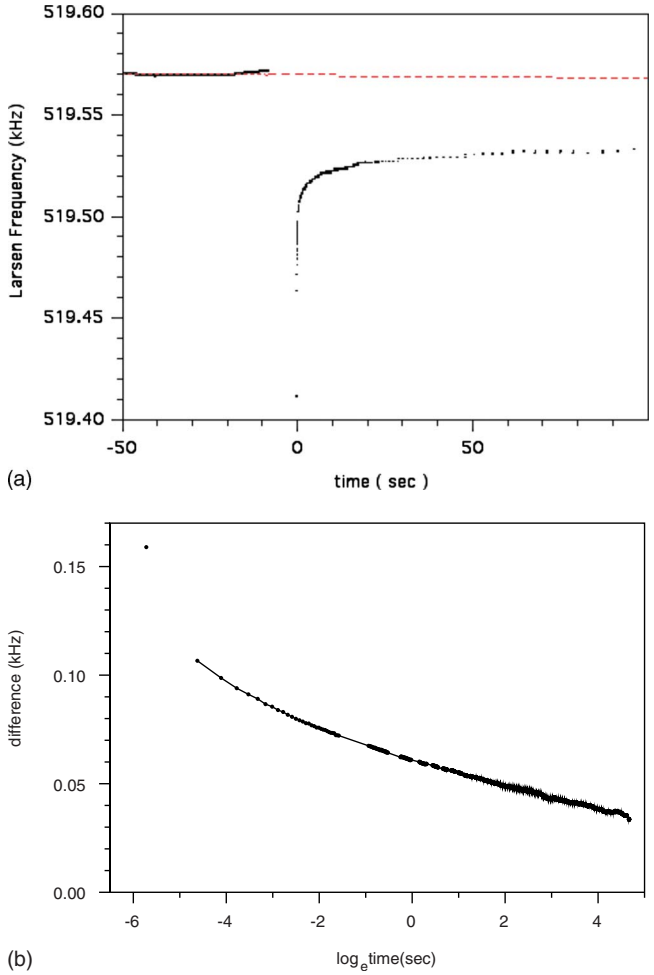


FIG. 11. (Color online) (a) Example of the response to a wooden ball impact on a sandstone sample. (b) $\log t$ relaxation, with accelerated recovery at early times, is also seen in the sandstone sample.

the frequency drops by about 25 Hz. The frequency then tends to recover. At first the recovery is fast; it then slows. It is not obvious whether recovery continues beyond the 400 s of study.

The recovery may be isolated by taking the difference between the instantaneous Larsen frequency and an extrapolation of the Larsen frequency from early times before the impact, thus removing the effects of temperature drift (based on a presumption that the temperature drift proceeds at a constant rate). The difference is plotted in Fig. 10(b) versus the log of time since the impact. As observed in other studies,[3–10] the change is linear in $\log t$. Previous work has been confined to times between 1 and 1000 s after the load is removed. Here the time dependence is extended back to milliseconds. It is seen that the $\log t$ behavior continues, at least in this sample, back to 22 ms. For earlier times yet (the first few data points pertain respectively to $t=3.3$ ms, 9.8 ms, and 16.4 ms), the relaxation is faster than $\log t$. Similar behaviors are observed in sandstone (Fig. 11).

It could be speculated that the accelerated relaxation at short times is an artifact of the signal processing or of the dynamics of the Larsen signal itself. It has not been observed in prior work.[3–10] Our studies (Figs. 5 and 7) of the dynamics of the Larsen frequency subsequent to a sudden change in circuit impedance show, however, that the Larsen dynamics itself has no time scales slower than an absorption time, here of the order of hundreds of microseconds. The accelerated relaxation is therefore attributed to the mechanics of the damaged cement or sandstone.

After 400 s, Fig. 10 shows a recovery of 80%. Extrapolation suggests that complete recovery, if it were to be achieved, would require another 6 h. A reliable study over such long times would require precise control of temperature over that period and has not been attempted here.

The above tests were repeated for three different ball sizes (diameter=3.2, 4.75, and 6.35 mm) and a variety of positions on the cement sample. In every case, the Larsen frequency was found to suddenly diminish at the time of impact and then recover slowly. Curves such as those in Figs. 10 and 11 are typical, but the magnitude of the drop was a random function of position. Figure 12 shows the slopes $\beta = df/d \ln t$, the Δf's at 3.27 ms (the soonest after the impact that we measure), and the Δf's at 1 s (in all cases well into the $\log t$ regime). Each data point is based on impacts at $N=10-20$ different positions. One sigma error bars are plotted; they were constructed by dividing observed standard deviations by $[N-1]^{1/2}$.

It is remarkable how well the data fit simple power law dependence on ball size. The rate of recovery $\beta=df/d \ln t$ scales very closely with the second power of ball radius. The value of Δf at (an arbitrarily chosen) time $t=1$ s within the linear regime scales roughly as the $3d$ power of ball radius. These two powers taken together indicate that the frequency shift at any fixed time is not going to be a simple power of ball radius, so this latter observation cannot be considered robust.

Figure 13 shows a time, $T_{\text{recovery}}=\exp(\Delta f(1.0\ \text{s})/\beta)$, at which recovery would be complete if the linear in $\log t$ behavior were to continue. In accord with the exponents seen for Δf at 1 s and β, one would expect T_{recovery} to be exponential in r. Such an r-dependence is not ruled out by the data which fit tolerably to 2 s times $\exp(1.1r/\text{mm})$, but $T_{\text{recovery}} \sim r^5$ is better supported. The smallness of the error bars shows that there is a positive, but not perfect, correlation between Δf and β across the several ball drops at fixed ball size. Otherwise the fractional error in T_{recovery} would be larger than that of both Δf and β.

The data also permit a calculation (not shown) of what Δf would be if we extrapolated the linear regime to time $t=3.27$ ms. The ratio of that Δf to the actual Δf at that time is very nearly independent of ball size and of ball impact position. In this sense the curvature seen in Fig. 10 is independent of these parameters. The average ratio $\Delta f(3.27\ \text{ms})/(\Delta f(1.0\ \text{s})+5.72\beta)$ was 1.51, 1.43, and 1.43 for the large, medium, and small balls, respectively. This measure of curvature thus appears to be not a function of the damaged volume. This ratio fluctuated randomly with impact position but only slightly; standard deviations were only 0.12, 0.14, and 0.175, respectively. Thus the curvature is also not a function of the position-dependent sensitivity of the Larsen frequency to damage.

It is tempting to conjecture as to the mechanics behind these exponents. That the frequency shifts at times of order of 1 s are proportional to r^3 suggests that it is a measure of

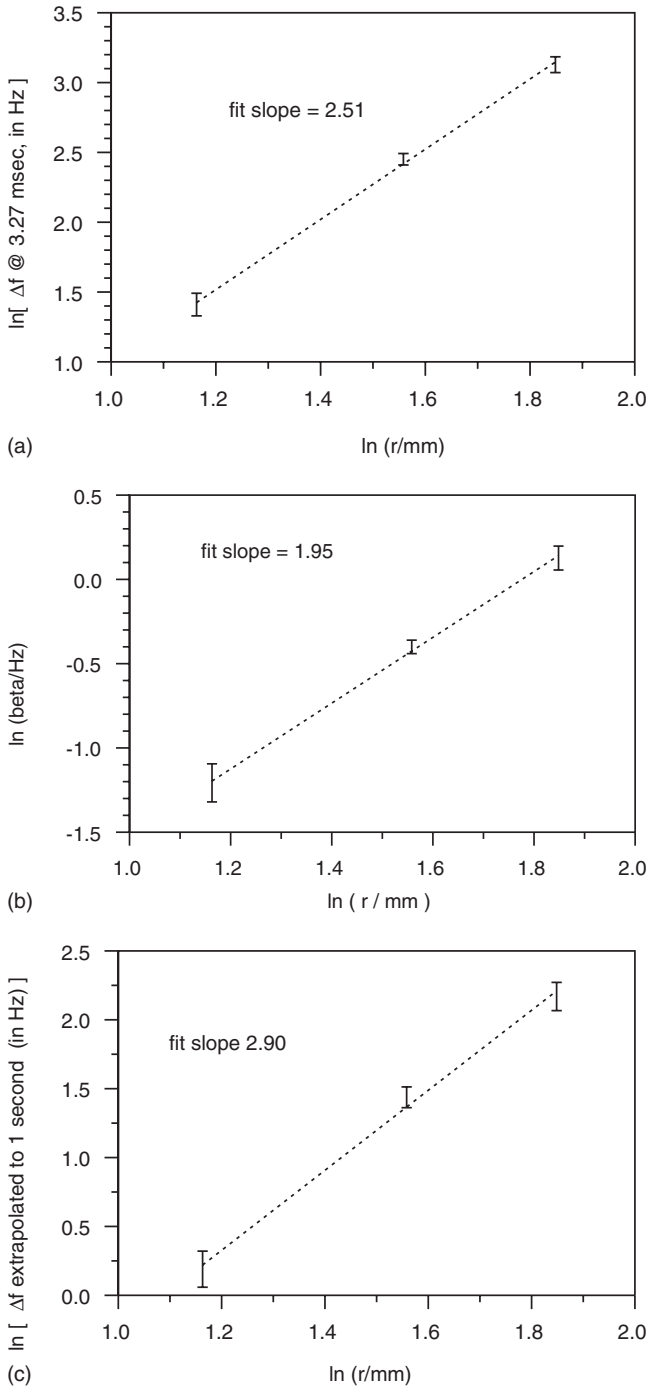


FIG. 12. (a) The shift in Larsen frequency from a base of about 500 kHz, as measured at the first accessible time of 3.27 ms, varies with the size of the wooden ball like $r^{2.5}$. (b) The slope $\beta = df/d \ln t$ varies with ball size like r^{-2}. (c) The frequency shift at a moderate time of 1 s varies like r^3.

the volume of damaged material. That the dependence is not stronger is consistent with the Hertzian elastic contact theory estimate that maximum stress is independent of ball size. That recovery rates scale with r^2 suggests maybe that recovery depends on the surface area of the damaged region. It would be interesting to repeat measurements like these with balls of different materials and/or different impact velocities. Inasmuch as the aim of this report is to illustrate the potential applications of Larsen monitoring, rather than to do materials testing, such is relegated to another occasion.

It is also tempting to compare the slow dynamics seen here with that reported extensively by the group at LANL. Their measurements are very different, using lower strains for the damage, but doing so over larger volumes. Their samples are loaded globally by low strain (of order 10^{-6}) continuous harmonic vibrations at ~3000 Hz. They observe $\Delta f/f$ values at time 1 s that are comparable to those observed here and that scale with strain. They observe a strain-independent time to recovery. Our (compressive) strains are nearly 10^4 times greater, but the volume we damage is only a part in 10^6 of the total volume. In spite of this enormous difference in the parameters of the nonlinear loading, the resulting $\Delta f/f$ and total time to recovery are similar. Precise comparisons would not be meaningful, as different materials can have much stronger or weaker slow dynamics; detailed comparison would require use of matched samples.

That the Larsen frequency always diminishes at impact is noteworthy. Similar behavior is seen in the LANL work.[3–10] There it is perhaps unsurprising, as a diminishment in modulus due to damage will—based on first-order perturbation theory—correspondingly diminish each of the isolated eigenfrequencies of vibration. Theory does not permit us to make a similar inference here. One imagines that diminishment of local modulus at the point of impact ought to lead to randomly signed changes in Larsen frequency, as does an addition of mass (Sec. III C). We are perplexed, but the measurements are unequivocal. Indeed, the frequency shifts associated with the different impact sites were well clustered around their mean, with standard deviations of only 30%.

V. CONCLUSION

Larsen monitoring may be useful for materials characterization or structural health monitoring. For this latter application, it may be compared to that of Lu and Michaels.[17] Its ability to detect small fast mechanical changes lends itself to the monitoring of cure evolution or damage accumulation, especially if these proceed discontinuously. This ability also lends itself to studies of mesoscale nonlinear elastic slow dynamics that would be complementary to those in the literature.

Before any significant application, however, there remain fundamental questions about Larsen signals in reverberant structures. Notable among these is our lack of understanding of the stability of a Larsen state against small perturbations and the rate and manner in which it relaxes; theory developed here does not fully agree with measurements. Nevertheless, some theoretical ideas are confirmed. The Larsen circuit tends to pick the most stable of candidate frequencies at which Im Gg is zero and/or the frequency with the largest Green's function. It may be that these matters can be better understood with numerical simulations and further experiments informed by numerical simulations and theory. Nonlinear dynamical systems composed of reverberant linear waves with electronic gain limited by simple saturation offer a rich topic for investigation.

ACKNOWLEDGMENTS

This work was supported by the NSF Grant No. CMS 05-28096. The authors thank Dr. Konstantine Maslov for

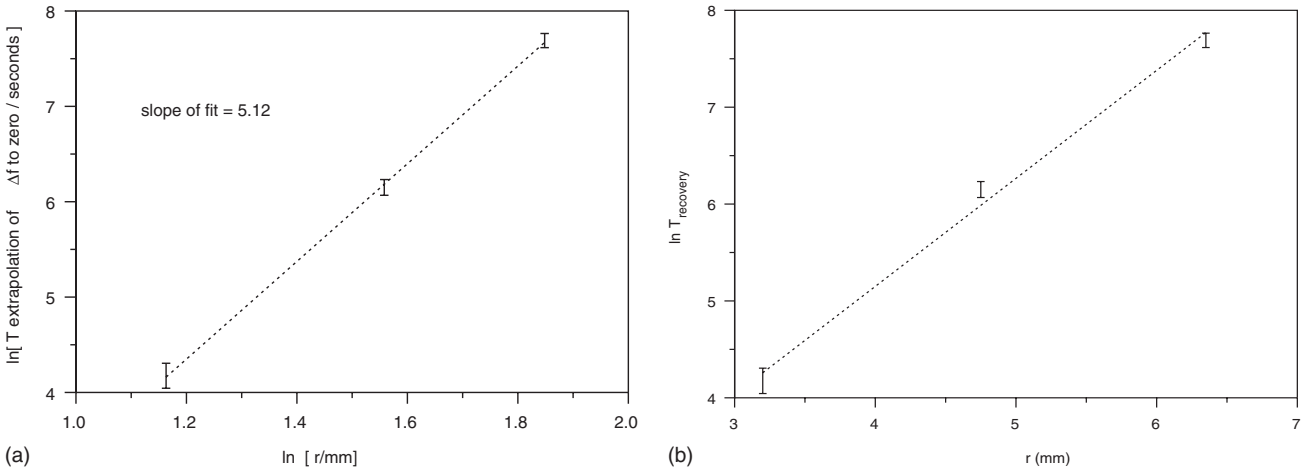


FIG. 13. The extrapolated time at which recovery would be complete were it to continue varies like r^5. An exponential dependence fits the data less well.

fruitful discussions on electronics. We thank John Popovics for the cement paste samples, and advice.

APPENDIX: DERIVATION OF EQUATION (7)

The left and right sides of Eq. (7) are obtained from Eq. (5) by distinct arguments. We start with the definition, Eq. (6),

$$S(t) = V(t)\exp(i\Omega t) + \text{c.c.} = 2|V(t)|\cos(\Omega t - \phi(t)), \quad \text{(A1)}$$

where we have written $V(t)=|V|\exp(-i\phi)$, $|V|$ and ϕ being real and slowly varying with time.

As S is periodic, with period $2\pi/\Omega$, with slowly varying amplitude and phase, so is the right side of Eq. (5), $f(S(t))$. Any periodic function $B(t)$ can be written as a complex Fourier series:

$$B(t) = \sum_{n=-\infty}^{+\infty} b_n \exp(in\Omega t), \quad \text{(A2)}$$

where the complex coefficients are given by an integral over any contiguous period $2\pi/\Omega$ in time,

$$b_n = (\Omega/2\pi)\int B(t)\exp(-in\Omega t)dt. \quad \text{(A3)}$$

We are presuming that S is essentially harmonic, with a single frequency Ω. $f(S(t))$ is not harmonic, but it is periodic. We confine attention to the primary components, $n=\pm 1$, of $f(S(t))$ by arguing that the transfer function gG is weak at the high frequency overtones. Thus

$$f(S(t)) = b_1 \exp(i\Omega t) + \text{c.c.}, \quad \text{(A4)}$$

where

$$b_1 = (\Omega/2\pi)\int f(2|V|\cos(\Omega t - \phi))\exp(-i\Omega t)dt. \quad \text{(A5)}$$

By defining $\Omega t=\zeta$ and $\tau=\zeta-\phi$, this becomes

$$b_1 = (1/2\pi)\int f(2|V|\cos(\zeta - \phi))\exp(-i\zeta)d\zeta$$

$$= (\exp(-i\phi)/2\pi)\int f(2|V|\cos(\tau))\exp(-i\tau)d\tau. \quad \text{(A6)}$$

We recall $V(t)=|V|\exp(-i\phi)$ and see that b_1 is $q(V_o)V(t)$, where q was defined in Eq. (8). Thus we establish the right side of Eq. (7) as equivalent to $f(S(t))$.

For the left side of Eq. (5), which is a linear temporal convolution, we write

$$H \otimes S = \int H(t-\tau)(V(\tau)\exp(i\Omega\tau) + \text{c.c.})d\tau. \quad \text{(A7)}$$

The integral is over all τ. Let $\zeta=t-\tau$ and write

$$H \otimes S = \int H(\zeta)(V(t-\zeta)\exp(i\Omega t - i\Omega\zeta) + \text{c.c.})d\tau$$

$$= \exp(i\Omega t)\int H(\zeta)(V(t-\zeta)\ \exp(-i\Omega\zeta)d\tau + \text{c.c.} \quad \text{(A8)}$$

The slowly varying function $V(t-\zeta)$ is now expanded in a Taylor series around t:

$$H \otimes S = \exp(i\Omega t)$$

$$\times \int H(\zeta)(V(t) - \zeta V'(t) + \cdots)\exp(-i\Omega\zeta)d\zeta + \text{c.c.}$$

$$= \exp(i\Omega t)\left[V(t)\int H(\zeta)\exp(-i\Omega\zeta)d\zeta\right.$$

$$\left. - V'(t)\int H(\zeta)\zeta\exp(-i\Omega\zeta)d\zeta + \cdots\right] + \text{c.c.}$$

$$= \exp(i\Omega t)[V(t)\tilde{H}(\Omega) - iV'(t)\tilde{H}'(\Omega) + \cdots] + \text{c.c.}$$

$$= \exp(i\Omega t)[\tilde{H}(\Omega - id/dt)V(t)] + \text{c.c.} \quad \text{(A9)}$$

This is the left side of Eq. (7).

[1]A. Larsen, "Ein akustischer Wechselstromerzeuger mit regulierbare Periodenzahl für schwache Ströme (An acoustic alternating-current generator

for weak currents, with adjustable frequency)," Elektrotech. Z., ETZ **32**, 284–285 (1911); A. Kjerbye Nielsen, "Larsen-effekten' og den fØrste elektriske tone generator baseret herpa(The 'Larsen effect' and the first electrical pure-tone generator based on this effect)," Teletek. (Engl. ed.) **3**, 140–148 (1984); D. Barbaro, "Self starting acoustic oscillations in closed spaces," Alta Freq. **27**, 472–485 (1958).
[2]R. L. Weaver and O. I. Lobkis, "On the line width of the ultrasonic Larsen effect in a reverberant body," J. Acoust. Soc. Am. **120**, 102–109 (2006).
[3]J. A. TenCate, E. Smith, and R. A. Guyer, "Universal slow dynamics in granular solids," Phys. Rev. Lett. **85**, 1020–1023 (2000).
[4]K. E. A. Van Den Abeele, A. Sutin, J. Carmeliet, and P. A. Johnson, "Micro-damage diagnostics using nonlinear elastic wave spectroscopy (NEWS)," NDT & E Int. **34**, 239–248 (2001).
[5]P. Johnson and A. Sutin, "Slow dynamics and anomalous fast dynamics in diverse solids," J. Acoust. Soc. Am. **117**, 124–130 (2005).
[6]J. A. TenCate, J. Duran, and T. J. Shankland, "Nonlinearity and slow dynamics in rocks: Response to changes of temperature and humidity," in *Nonlinear Acoustics at the Beginning of the 21st Century*, edited by O. V. Rudenko and O. A. Sapozhnikov (MSU, Moscow, 2002), pp. 767-770.
[7]J.-C. Lacouture, P. A. Johnson, and F. Cohen-Tenoudji, "Study of critical behavior in concrete during curing by application of dynamic linear and nonlinear means," J. Acoust. Soc. Am. **113**, 1325–1332 (2003).
[8]O. O. Vakhnenko, V. O. Vakhnenko, T. J. Shankland, and J. TenCate, "Strain induced kinetics of inter-grain defects as the mechanism of slow dynamics in the nonlinear resonant response of humid sandstone bars," Phys. Rev. E **70**, 015602 (2004).
[9]O. O. Vakhnenko, V. O. Vakhenko, and T.-J. Shankland, "Soft-ratchet of end-point memory in the nonlinear resonant response of sedimentary rocks," Phys. Rev. B **71**, 174103 (2005).
[10]R. A. Guyer and P. A. Johnson, "Nonlinear mesoscopic elasticity: Evidence for a new class of materials," Phys. Today **52**, 30–35 (1999).
[11]U. Kaatze, F. Eggers, and K. Lautscham, "Ultrasonic velocity measurements in liquids with high resolution techniques, selected applications and perspectives," Meas. Sci. Technol. **19**, 062001 (2008); S. Garrett, "Resonant acoustic determination of elastic moduli," J. Acoust. Soc. Am. **88**, 210–221 (1990); E. J. Chern, J. S. Heyman, and J. H. Cantrell, Jr., "Determination of material stress from the temperature dependence of the acoustic natural velocity," Ultrasonics Symposium, pp. 960–963 (1981).
[12]Our comparator is an electronic device which switches its output between $\pm F_o$ when its input changes sign.
[13]C. Pierre, "Comments on Rayleigh's quotient and perturbation theory for the eigenvalue problem," ASME J. Appl. Mech. **55**, 986–988 (1988).
[14]T. Ericson, "A theory of fluctuations in nuclear cross sections," Ann. Phys. (N.Y.) **23**, 390–414 (1963); R. H. Lyon, "Statistical analysis of power injection and response in structures and rooms," J. Acoust. Soc. Am. **45**, 545-565 (1969).
[15]K. L. Johnson, *Contact Mechanics* (Cambridge University Press, Cambridge, England, 1985).
[16]The impacts are clearly inelastic (coefficients of restitution are less than 40%) so the theory (Ref. 15) based on linear elastostatics must be treated as approximate. Interestingly, coefficients of restitution are about the same when the wooden balls impact on steel; most of the energy loss is in the wood.
[17]Y. Lu and J. E. Michaels, "A methodology for structural health monitoring with diffuse ultrasonic waves in the presence of temperature variations," Ultrasonics **43**, 717–731 (2005).

Wideband nonlinear time reversal seismo-acoustic method for landmine detection

Alexander Sutin
Stevens Institute of Technology, Hoboken, New Jersey 07030 and ARTANN Laboratories, Inc., 1459 Lower Ferry Road, Trenton, New Jersey 08618

Brad Libbey
U.S. Army RDECOM, CERDEC, Night Vision and Electronics Sensors Directorate, 10221 Burbeck Road, Fort Belvoir, Virginia 22060

Laurent Fillinger
Stevens Institute of Technology, Hoboken, New Jersey 07030

Armen Sarvazyan
ARTANN Laboratories, Inc., 1459 Lower Ferry Road, Trenton, New Jersey 08618

(Received 10 June 2007; revised 31 December 2008; accepted 26 January 2009)

Acoustic and seismic waves provide a method to localize compliant mines by vibrating the top plate and a thin soil layer above the mine. This vibration is mostly linear, but also includes a small nonlinear deviation. The main goal of this paper is to introduce a method of processing that uses phase-inversion to observe nonlinear effects in a wide frequency band. The method extracts a nonlinear part of surface velocity from two similar broadcast signals of opposite sign by summing and canceling the linear components and leaving the nonlinear components. This phase-inversion method is combined with time reversal focusing to provide increased seismic vibration and enhance the nonlinear effect. The experiments used six loudspeakers in a wood box placed over sand in which inert landmines were buried. The nonlinear surface velocity of the sand with a mine compared to the sand without a mine was greater as compared to a linear technique.
 [DOI: 10.1121/1.3083223]

PACS number(s): 43.25.Ts, 43.40.Ga, 43.60.Tj [KA] Pages: 1906–1910

I. INTRODUCTION

The Army seeks methods of detecting landmines using acoustic and seismic wave energy. Most landmines are compliant objects that vibrate readily when subjected to a fluctuating pressure or stress field produced by an external source and this motion induces surface vibrations above a mine that are larger than in the surrounding soil.[1–4] The spatial extent of high vibration above the mine is generally less than a wavelength in soil in a detection band of 50–500 Hz. At higher frequencies, surface waves backscattered off the buried mine have been observed.[5–7] In earlier acoustic detection methods, the surface vibration was assumed to be proportional to the excitation force and nonlinear effects connected with soil and mine vibrations were not taken into account.

Nonlinear acoustic effects can be significant for seismic waves[8,9] and for waves in granular media.[10,11] Previous experiments suggest that nonlinear effects are pronounced in the presence of mines.[12–17] The most investigated nonlinear effect is generation of combination frequency vibrations under simultaneous excitation with two frequency waves.[12–15] This method works effectively for excitation frequencies close to the mine resonance, but is less effective at other frequencies because the nonlinear mine response strongly depends on the linear vibration amplitude, which is low at frequencies away from resonances. Consequently, tests with stepped excitation frequencies are required for detection to guarantee that the optimal frequency is excited since a mine's resonances depend on the type of mine, soil, and burial depth.[1,2,18] An alternate approach is to first find the resonant frequency using linear techniques.[1–4] Both of these methods require significant time for measurements, which makes them less practical for field application than wide band methods.

Time-reversed acoustic methods are used in this work because they provide larger amplitude soil vibrations than unfocused acoustic excitation[19–24] and facilitate measurements of nonlinear responses in the mine and soil. This method leads to large vibrations by generating signals with appropriate temporal delays, ensuring that the seismic and airborne arrivals are in phase in a specific soil region. The resulting pulse vibrates a mine over a wide frequency band.

The time reversal method reduces undesired nonlinear components associated with loudspeakers and propagation, while maintaining high amplitudes at the region of interest. Because the time reversal method focuses the energy spatially, lower source drive levels are required to generate ground vibration of the same amplitude as unfocused excitation systems; this decreases the nonlinear components generated by the speaker system. The spatial focusing maintains the high amplitude of seismic vibration and associated nonlinear effects in the focal area while reducing nonlinearities associated with the soil propagation path.

 0001-4966/2009/125(4)/1906/5/$25.00

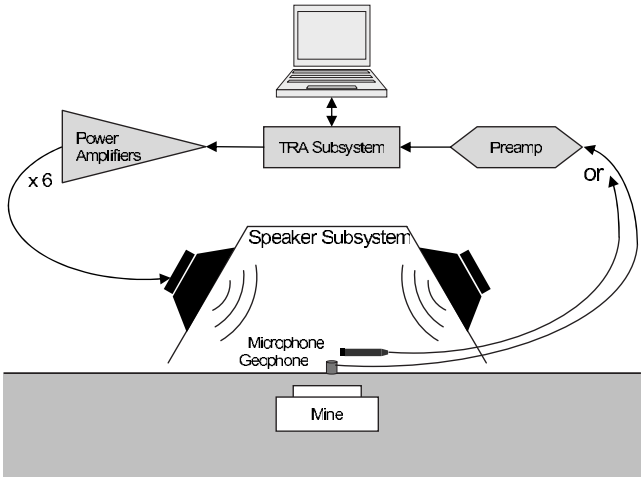


FIG. 1. The schema of the experimental setup for nonlinear time reversal landmine detection. Six simultaneous loudspeaker channels are synchronized with one acquisition channel.

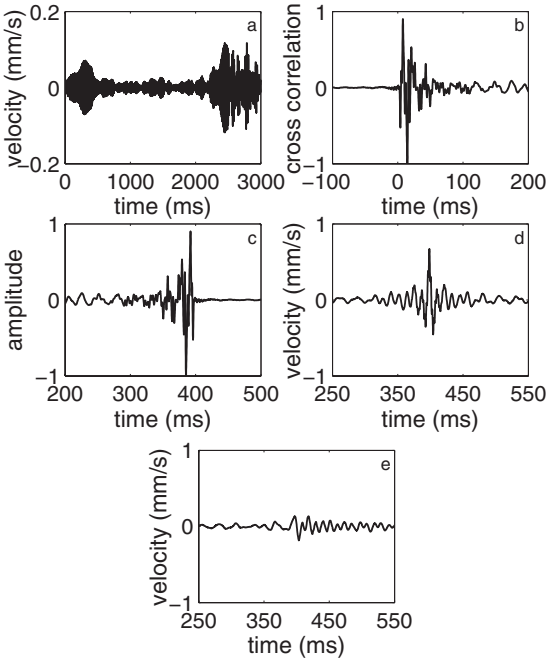


FIG. 2. Examples of signals used during the experiments: (a) surface vibration recorded by a geophone and corresponding to the first loudspeaker, (b) system impulse response estimate based on the cross-correlation of sweep signal and recorded vibration, (c) the time-reversed signal prepared for subsequent broadcasts, (d) focused vibration measurement resulting from the simultaneous time-reversed radiation from all loudspeakers and measured at the focus point, and (e) similar to (d) except measured at 0.6 m from the focus.

This paper investigates the phase-inversion method to measure nonlinear mine vibrations over a broad frequency range. Two similar broadband signal sets with opposite signs are broadcast sequentially and the two recorded responses are summed in postprocessing to cancel the linear component, leaving the nonlinear component. The phase-inversion method was first proposed for ultrasonic imaging of microbubble contrast agents in medical ultrasound. Bubbles have nonlinear responses several orders of magnitude greater than that of tissue and are detectable by considering the response at harmonics or combination frequencies,[25,26] similar to mines. Phase-inversion shows more contrast for bubble detection than the harmonic or combination frequency method and it is widely used in cardiac ultrasonic diagnostics.[25–28]

II. EXPERIMENTAL METHOD

The key components of the experimental setup are the electronic signal control subsystem developed at Artann Laboratories and a speaker subsystem developed at Stevens Institute of Technology[15] (Fig. 1). The electronic subsystem synchronizes six output channels that drive the six speakers and one input channel used for acquisition. A computer controls this subsystem by using custom software to generate, record, process, and store the signals. The speaker subsystem provides an enclosed acoustic space where multiple reflections lead to better focusing of the acoustic signals. The system consists of six speakers, Sony XS-L1265G, with a diameter of 30 cm mounted on the walls of a box in the shape of a hexagonal prism. The vertex of each speaker cone was 0.3 m above the soil surface. The box had a closed top and open base, 1 m across, such that the acoustic field within the enclosure excites the sand surface. The box was hung from a wheeled cart so the lowest edge was 3 cm above the soil.

The test was conducted at an Army test facility in a sand box 1.2 m deep, 3.0 m wide, and 32.7 m long. The *P*-wave and *S*-wave speeds were previously measured and found to be 151 and 118 m/s, respectively. The white beach sand's mineralogy was 98% quartz, 1.5% illite and mica, 0.4% kaolinite, and 0.1% chlorite, with a combined specific gravity of 2.65. The median grain size by mass is 0.3 mm, and the distribution of particle sizes shows 95% of the mass having a grain size between 0.1 and 0.6 mm and 80% having a grain size between 0.2 and 0.5 mm. The current data were collected with soil temperatures of 24–28 °C.

The surface vibrations were measured using an array of geophones without coupling spikes placed on the soil surface. Fifteen geophones, IO SM-11, were available for spatial sampling at 3 in. increments. The mass of a geophone is much less than the mass of soil above the mine, but can lead to a small variation in the resonant frequency.[16,29] A $\frac{1}{2}$ in. microphone, BK 11.5 mV/Pa, was positioned near the focus point 2 cm above the surface in order to compare system nonlinearities to those generated in the sand.

The time reversal focusing was conducted to provide a large amplitude seismic displacement over a large bandwidth. Details of the time reversal multistep process have been previously described[22] and will be briefly recounted here. In the first step, a sweep frequency signal, with frequency from 500 to 50 Hz, was broadcast from one loudspeaker. The resulting seismic vibration was recorded from one geophone at the focal point near the center of the loudspeaker box [Fig. 2(a)]. Next, the cross-correlation was calculated between the recorded vibration and the initial sweep frequency signal. This cross-correlation is proportional to the system impulse response measured in the frequency band 50–500 Hz. The calculated impulse response was normalized by setting the maximum amplitude to 1 [Fig. 2(b)]. Finally, the time reversal occurred by flipping the normalized

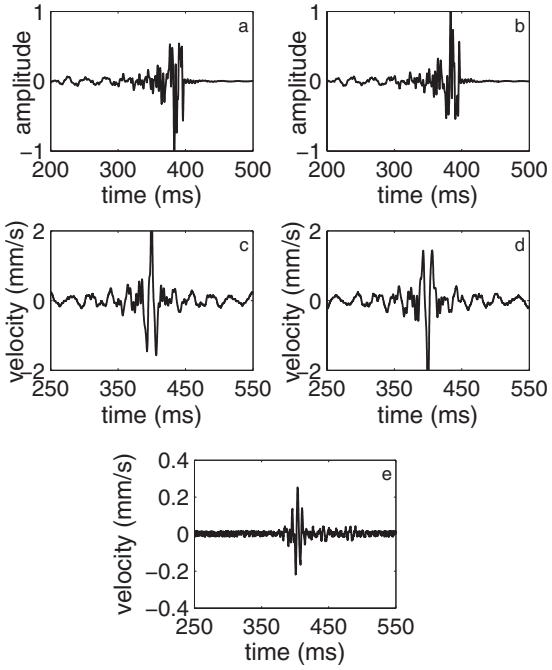


FIG. 3. Representative signals used in the phase-inversion method and measured above a VS-1.6 landmine buried at 5 cm: (a) one channel of the time-reversed impulse response set, (b) similar to (a) except with opposite sign, (c) and (d) reversed and focused signals as measured by the geophone for the standard and phase inverted sets, and (e) the nonlinear response, a sum of the two phase inverted signals divided by 2.

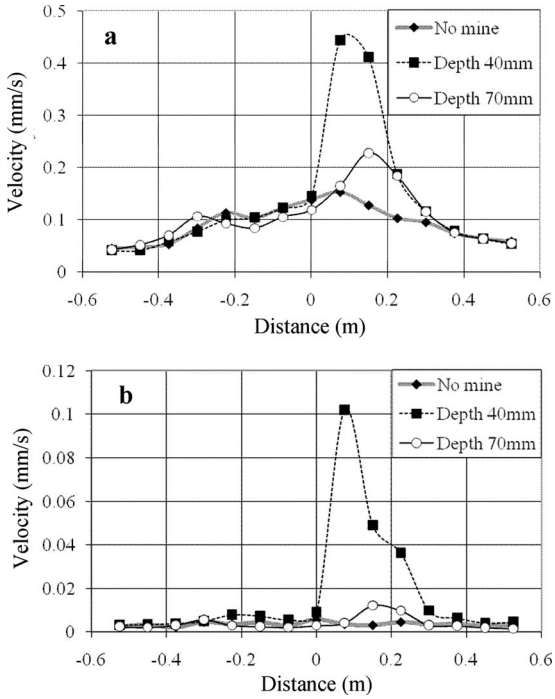


FIG. 4. The peak velocity of the sand surface measured at points on a line passing through the time-reversed focus location. (a) The linear response where the presence of a VS-1.6 antitank landmine increases the local amplitude. (b) The nonlinear response shows a greater mine/no-mine contrast than the linear response.

impulse response [Fig. 2(c)]. This sequence of steps was repeated for all six channels creating a set of six time-reversed impulse responses.

This set was broadcast simultaneously from the loudspeakers. The result was a physical match filter for each channel in the environment that focuses the seismic energy near the selected geophone when broadcast [Fig. 2(d)]. The amplitude of surface vibration decreased with the distance from the focus point when the simultaneous broadcast was measured at different locations [Fig. 2(e)]. The benefit of time reversal processing as used here is its ability to produce large amplitude seismic disturbances that increase the nonlinear response.

III. NONLINEAR TIME REVERSAL TEST WITH A LANDMINE

Time reversal focusing creates a short vibration pulse in the sand that is ideal for the nonlinear phase-inversion method. The radiated signals described above are broadcast twice but with opposite signs, one followed by the other [Figs. 3(a) and 3(b)]. The received signals consist of linear and nonlinear components [Figs. 3(c) and 3(d)]. The linear components have opposite signs, as is apparent in the figure. However, even ordered nonlinear components of these two signals have the same sign, and when the two signals are added together the linear components cancel each other and the even ordered nonlinear components add constructively. The summations are twice the nonlinear amplitude generated in a single pulse so the responses of the summed signals are divided by 2 [Fig. 3(e)]. This summation procedure also eliminates odd ordered nonlinearities, which played a lesser role in earlier experiments.[12,15,16]

Preliminary data from the microphone and geophone were compared in sand without a mine using the phase-inversion process to be sure that nonlinearities are generated in the sand as opposed to the loudspeakers. The pressure ratio of the peak amplitude of the nonlinear component to the peak amplitude of the combined linear and nonlinear measured signals is 7.5×10^{-3} as measured on the microphones and the same ratio for soil velocities is 86×10^{-3} as measured on the geophone. This indicates that the majority of the nonlinear component is being generated in the soil and relatively little is generated by the equipment or in the air.

The principle testing was conducted over sand with and without landmine using the geophone array. Three mines were tested individually: an Italian VS-1.6 antitank with a diameter of 22 cm, a Russian PMN antipersonnel with a diameter of 11 cm, and an Italian VS-50 antipersonnel with a diameter of 9 cm. Surface velocity time traces were measured at locations on a line using the techniques described above. The peak linear velocity of these traces illustrates that the response is greater above the VS-1.6 and PMN mines than in the sand without a mine, but not so with the VS-50 mine, Figs. 4(a) and 5(a). However, the peak nonlinear velocity is greater than the sand without a mine for all of the mines tested. The mine/no-mine contrast ratio is an important detection parameter and is computed as a ratio of the maximum peak velocity over the mine to the peak velocity without the mine at the same point. This ratio is greater for the nonlinear responses than the linear, Table I.

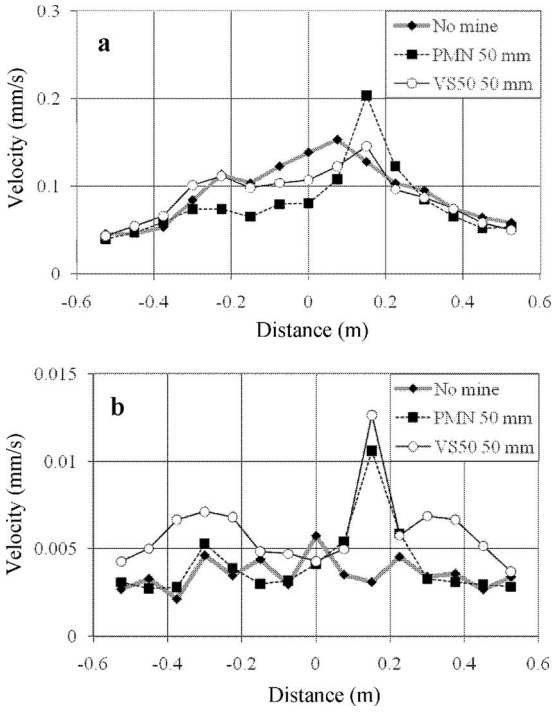


FIG. 5. The peak velocity of the sand surface measured at points on a line passing through the time-reversed focus location. (a) The linear response where the presence of antipersonnel landmines increases the local amplitude. (b) The nonlinear response shows a greater mine/no-mine contrast than the linear response.

IV. CONCLUSION

The nonlinear effects connected with mine vibration in sand were investigated over a wide frequency band using the phase-inversion method. Time reversal acoustic focusing excited the soil/mine vibration by increasing seismic energy in the region being interrogated. A mine/no-mine contrast was observed for linear and nonlinear responses. The contrast for the nonlinear response was higher than for linear response.

This paper demonstrates that phase-inversion processing of a pair of wide band pulses, computed using time reversal techniques, is an effective tool for measuring mine nonlinearities. The technique could be implemented on many sensor systems being developed for landmine detection. These systems often combine spatial sampling of surface vibration using laser vibrometers or contact sensors with acoustic or vibration excitation.[1–4] The implementation of the nonlinear time reversal technique to these systems was demonstrated in a recent test where time reversal focusing was used with the landmine detection system that includes four shakers and a 16-beam laser Doppler vibrometer for fast scanning of a soil region.[4] Tests of this system in the time reversal regime demonstrated good spatial and temporal focusing in soil without a mine. This suggests that broadband nonlinear data could be obtained using existing linear mine detection systems, ultimately improving landmine detection.

TABLE I. Surface velocity contrast ratios.

Mine type	Depth (mm)	Contrast ratio	
		Linear	Nonlinear
VS-1.6 antitank	40	2.9	28.9
VS-1.6 antitank	70	1.8	3.9
PMN antipersonnel	50	1.6	3.4
VS-50 antipersonnel	50	1.1	4.0

ACKNOWLEDGMENTS

The U.S. Army, RDECOM, CERDEC Night Vision and Electronic Sensors Directorate supported this work under Contract No. W909MY-05-C-0007. The authors would like to acknowledge Stevens Institute of Technology for providing the six-channel speaker enclosure.

[1]N. Xiang and J. M. Sabatier, "An experimental study on antipersonnel landmine detection using acoustic-to-seismic coupling," J. Acoust. Soc. Am. **113**, 1333–1341 (2003).
[2]J. M. Sabatier, "Land mine detection measurements using acoustic-to-seismic coupling," Proc. SPIE **4038**, 645–655 (2000).
[3]V. Valeau, J. Sabatier, R. D. Costley, and N. Xiang, "Development of a time-frequency representation for acoustic detection of buried objects," J. Acoust. Soc. Am. **116**, 2984–2995 (2004).
[4]V. Aranchuk, J. M. Sabatier, A. K. Lal, C. F. Hess, R. D. Burgett, and M. O'Neill, "Multi-beam laser Doppler vibrometry for acoustic landmine detection using airborne and mechanically-coupled vibration," Proc. SPIE **5794**, 624–631 (2005).
[5]A. Behboodian, W. R. Scott, Jr., and J. H. McClellan, "Signal processing of elastic surface waves for localizing buried land mines," Proceedings of the 33rd Assilomar Conference on Signal, Systems, and Computers, Assilomar, CA, October, 1999.
[6]W. R. Scott, Jr., G. D. Larson, J. S. Martin, and G. S. McCall II, "Field testing and development of a seismic landmine detection system," Proc. SPIE **5089**, 643–652 (2003).
[7]T. G. Muir, M. E. Zakharia, A. Gril, and E. G. de Garambe, "Landmine detection with seismic sonar," Proc. SPIE **5415**, 145–155 (2004).
[8]A. V. Nikolaev, "Problems of nonlinear seismology," Phys. Earth Planet. Inter. **50**, 1–7 (1988).
[9]O. V. Pavlenko, "Nonlinear seismic effects in soils: Numerical simulation and study," Bull. Seismol. Soc. Am. **91**, 381–396 (2001).
[10]V. Tournat, V. Yu. Zaitsev, V. E. Nazarov, V. É. Gusev, and B. Castagnéde, "Experimental study of nonlinear acoustic effects in a granular medium," Acoust. Phys. **51**, 543–553 (2005).
[11]P. A. Johnson and X. Jia, "Nonlinear dynamics, granular media and dynamic earthquake triggering," Nature (London) **473**, 871–874 (2005).
[12]D. M. Donskoy, A. Ekimov, N. Sedunov, and M. Tsionskiy, "Nonlinear seismo-acoustic land mine detection and discrimination," J. Acoust. Soc. Am. **111**, 2705–2714 (2002).
[13]D. M. Donskoy and A. M. Sutin, "Method and apparatus for acoustic detection of mines and other buried man-made objects," U.S. Patent No. 5,974,881 (2 November 1999).
[14]J. M. Sabatier, M. S. Korman, and N. Xiang, "Linear and nonlinear acoustic velocity profiles over buried land mines," Proc. SPIE **4742**, 695–700 (2002).
[15]D. M. Donskoy, A. Reznik, A. Zagrai, and A. Ekimov, "Nonlinear vibrations of buried landmines," J. Acoust. Soc. Am. **117**, 690–700 (2005).
[16]K. Attenborough, Q. Qin, J. Jefferis, and G. Heald, "Accelerometer measurements of acoustic-to-seismic coupling above buried objects," J. Acoust. Soc. Am. **122**, 3230–3241 (2007).
[17]M. S. Korman and J. M. Sabatier, "Nonlinear acoustic techniques for landmine detection," J. Acoust. Soc. Am. **116**, 3354–3369 (2004).
[18]W. C. K. Alberts II, J. M. Sabatier, and R. Waxler, "Resonance frequency shift saturation in land mine burial simulation experiments," J. Acoust. Soc. Am. **120**, 1881–1886 (2006).
[19]P. D. Norville and W. R. Scott, Jr., "Time-reversal focusing of elastic surface waves," J. Acoust. Soc. Am. **118**, 735–744 (2005).
[20]P. D. Norville and W. R. Scott, Jr., "Time-reversal focusing of elastic waves in inhomogeneous media: An application to an elastic-wave landmine detection system," Proc. SPIE **5794**, 632–672 (2005).
[21]A. Sutin, P. Johnson, J. TenCate, and A. Sarvazyan, "Time reversal

acousto-seismic method for land mine detection," Proc. SPIE **5794**, 706–716 (2005).
[22]A. Sutin, B. Libbey, V. Kurtenoks, D. Fenneman, and A. Sarvazyan, "Nonlinear detection of land mines using wide bandwidth time-reversal techniques," Proc. SPIE **6217**, 398–409 (2006).
[23]B. Libbey and D. Fenneman, "Acoustic to seismic ground excitation using time reversal," Proc. SPIE **5794**, 643–654 (2005).
[24]P. D. Norville and W. R. Scott, Jr., "Time-reversal focusing of elastic surface waves with an asymmetric surface layer," J. Acoust. Soc. Am. **122**, EL95–EL100 (2007).
[25]S. Krishnan and M. O'Donnell, "Transmit aperture processing for nonlinear contrast agent imaging," Ultrason. Imaging **18**, 77–105 (1996).
[26]B. B. Goldberg, "Contrast agents," Ultrasound Med. Biol. **26**, S33–S34 (2000).
[27]P. J. A. Frinking, A. Bouakaz, J. Kirkhorn, F. J. TenCate, and N. de Jong, "Ultrasound contrast imaging: Current and new potential methods," Ultrasound Med. Biol. **26**, 965–975 (2000).
[28]A. Bauer, P. Hauff, J. Lazenby, P. Von Behren, M. Zomack, M. Reinhardt, and R. Schlief, "Wideband harmonic imaging: A novel contrast ultrasound imaging technique," Eur. Radiol. **9**, S364–S367 (1999).
[29]J. S. Martin, G. D. Larson, and W. R. Scott, Jr., "Surface-contacting vibrometers for seismic landmine detection," Proc. SPIE **5794**, 590–600 (2005).

Finite difference time domain electroacoustic model for synthetic jet actuators including nonlinear flow resistance

Gerben Kooijman[a] and Okke Ouweltjes
Philips Research Laboratories, High Tech Campus 36, 5656AE Eindhoven, The Netherlands

(Received 11 July 2008; revised 16 January 2009; accepted 22 January 2009)

A lumped element electroacoustic model for a synthetic jet actuator is presented. The model includes the nonlinear flow resistance associated with flow separation and employs a finite difference scheme in the time domain. As opposed to more common analytical frequency domain electroacoustic models, in which the nonlinear resistance can only be considered as a constant, it allows the calculation of higher harmonics, i.e., distortion components, generated as a result of this nonlinear resistance. Model calculations for the time-averaged momentum flux of the synthetic jet as well as the radiated sound power spectrum are compared to experimental results for various configurations. It is shown that a significantly improved prediction of the momentum flux—and thus flow velocity—of the jet is obtained when including the nonlinear resistance. Here, the current model performs slightly better than an analytical model. For the power spectrum of radiated sound, a reasonable agreement is obtained when assuming a plausible slight asymmetry in the nonlinear resistance. However, results suggest that loudspeaker nonlinearities play a significant role as well in the generation of the first few higher harmonics.
 [DOI: 10.1121/1.3081514]

PACS number(s): 43.28.Ra, 43.25.Nm, 43.25.Ed, 43.38.Ja [AVH] Pages: 1911–1918

I. INTRODUCTION

Synthetic jets are used in various application fields such as aerodynamic flow control, propulsion, micro-pumps, surface cleaning, and heat transfer enhancement, see e.g., Refs. 1–5. For the purpose of cooling of electronic products by means of forced convection, synthetic jets are believed to be promising due to a number of advantages over fan-based technology:

- lower noise level;
- better efficiency: less power needed for same cooling performance;
- higher reliability, longer lifetime;
- lower cost;
- beneficial form factor allows for high degree of freedom in design and ease of installation in devices, also miniaturization is possible;
- ideal for spot-cooling: cooling of small components that generate a lot of heat, which are found ever more in electronic devices.

Note that some other techniques for localized cooling of electronics are discussed in, e.g., Ref. 6.

Research on efficient sound reproduction at low frequencies by a small form-factor device has recently led to the development of a special resonant loudspeaker system.[7] This system basically consists of a loudspeaker connected to a tube, see Fig. 1(a). It was found that this system could also be used as a synthetic jet actuator. The acoustic motion of the fluid results in the formation of a pulsating jet at the outlet of the tube: as air is pushed out of the tube during one-half of a cycle, flow separation at the edges of the outlet occurs (provided that the Strouhal number is low enough, see, e.g., Ref. 8), and a jet is formed. During the intake part of a cycle, air is sucked into the tube from all directions. Averaged over a complete cycle no mass is injected into the fluid domain. However, due to the described difference between the inward motion of air at intake and the outward motion in the form of a jet, a net momentum is injected into the fluid domain. Therefore, synthetic jet actuators are also known as zero-net mass flux, nonzero momentum flux devices.

To diminish the radiated sound, a dipole actuator, as shown in Fig. 1(b), has been proposed.[9] Here, basically, two tubes are placed on either side of a loudspeaker. In this way the acoustic motions at the outlets of the tubes are in antiphase, minimizing the generated sound. Similarly, a quadrupole synthetic jet cooler, comprising two loudspeakers and four tubes, has been presented.[9] As an alternative, in order to reduce the radiated sound, the enclosed backing volume of the original resonant loudspeaker system can be omitted,[10] see Fig. 1(c). This yields, due to the resulting acoustic shortcut between the back of the loudspeaker and the tube outlet, a geometrically asymmetric dipole actuator: the acoustic volume velocity at the back of the loudspeaker and at the tube outlet are (more or less) equal in magnitude but opposite in phase.

In this paper, a lumped element electroacoustic model of such an asymmetric dipole actuator is presented. The model includes the nonlinear resistance associated with the flow separation at the tube ends. Furthermore, the model employs a finite difference scheme in the time domain. This allows the calculation of the higher harmonics of the fundamental signal with which the actuator is driven, which are generated due to the nonlinear resistance. This is opposed to more common analytical frequency domain electroacoustic models in-

[a]Electronic mail: gerben.kooijman@philips.com

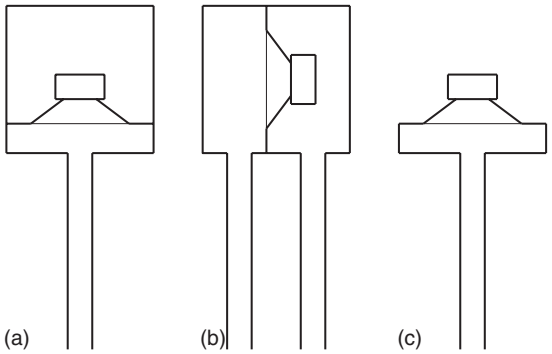


FIG. 1. Schematic drawing of (a) resonant loudspeaker system (Ref. 7), (b) dipole cooler (Ref. 9), and (c) asymmetric dipole cooler (Ref. 10).

cluding nonlinear flow resistance, see, e.g., Ref. 11. There, the variation in the nonlinear resistance with time (as it is proportional to the acoustic velocity) cannot be accounted for, and consequently no higher harmonics can be calculated.

Model predictions for the momentum flux of the synthetic jet will be compared to experimental results. Also, a measured power spectrum including generated higher harmonics of the driving signal will be compared to a model calculation. The latter is particularly of interest in the application of cooling as for these actuators acoustic annoyance is, besides tube resonances, mainly caused by higher harmonics, rather than by the low fundamental driving frequency.

II. MODEL

Lumped element modeling (see, e.g., Ref. 12) is employed for the asymmetric dipole synthetic jet actuator shown in Fig. 1(c). The equivalent circuit including the electrical and mechanical domain of the loudspeaker and the acoustical domain is depicted in Fig. 2. In the respective domains the voltage e, force F, and pressure p are the effort variables, and the current I, velocity v, and volumetric flow rate U (often called volume velocity in acoustics) are the flow variables. Coupling between the electrical and the mechanical domain is by a gyrator according to $F=BlI$ and $e_{\mathrm{emf}}=Blv$, with Bl the force factor and e_{emf} the counter-electromotive force. Coupling between the mechanical and the acoustical domains is by a transformer according to $p=F/S_d$ and $U=vS_d$, with S_d the loudspeaker's membrane area. The considered lumped elements are the electrical resistance R_e and inductance L_e of the loudspeaker's voice coil, and the mechanical resistance R_m, mechanical compliance C_m, and moving mass M_m of the loudspeaker's cone. Furthermore, in the acoustical domain we have the acoustical compliance C_{ac}:

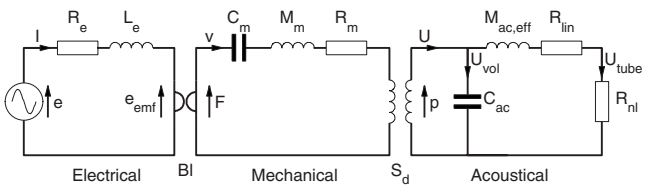


FIG. 2. Lumped element circuit representation of the asymmetric dipole synthetic jet actuator.

$$C_{\mathrm{ac}}=\frac{V}{\rho c^2}, \tag{1}$$

of the volume V between the loudspeaker and tube, with ρ and c the mass density, respectively, speed of sound. The tube can be considered as a lumped acoustical mass M_{ac}:

$$M_{\mathrm{ac}}=\frac{\rho L}{S}, \tag{2}$$

with L and S its length (including end correction) and cross sectional area, respectively. However, damping in wave propagation, which can be accounted for if the tube were modeled as a transmission line, is then not considered in first instance. Nevertheless, when approximating the transfer of acoustic pressure and volume velocity over a transmission line for low frequency: $(\omega L/c)^2 \ll 1$, it is found that the tube can be considered as an "effective" acoustical mass:

$$M_{\mathrm{ac,eff}}=\frac{\mathrm{Re}(k)}{k_0}M_{\mathrm{ac}}, \tag{3}$$

in series with a linear resistance:

$$R_{\mathrm{lin}}=-cM_{\mathrm{ac}}\,\mathrm{Im}(k). \tag{4}$$

The linear resistance accounts for the wave damping (for the fundamental frequency). Here k is the wavenumber, and $k_0=\omega/c$. Note that in the above the acoustic radiation impedance at the loudspeaker back and at the tube outlet is neglected. Accounting for visco-thermal losses at the tube walls, the wavenumber is given by[13]

$$k=-i\frac{\omega}{c}\sqrt{\gamma\frac{J_0(\sigma)}{J_2(\sigma)}+\gamma-1},$$

$$\sigma=0.72i^{3/2}\sqrt{Pr}Sh, \tag{5}$$

where J_0 and J_2 are Bessel functions of the first kind of zeroth and second order, respectively. Sh is the shear number: $Sh=(d/2)\sqrt{\omega/\nu}$, with d the diameter of the tube and ν the kinematic viscosity. Pr is the Prandtl number, and γ is the ratio of specific heats. Alternatively, for low Helmholtz numbers: $He=\omega d/(2c)\ll 1$, and high shear numbers: $Sh\gg 1$, the following expression can be used:[13,14]

$$k=\frac{\omega}{c}\left(1+\frac{1-i}{\sqrt{2}}\frac{1}{Sh}\left(1+\frac{\gamma-1}{\sqrt{Pr}}\right)-\frac{i}{Sh^2}\left(1+\frac{\gamma-1}{\sqrt{Pr}}-\frac{1}{2}\gamma\frac{\gamma-1}{Pr}\right)\right). \tag{6}$$

In the remainder Eq. (5) will be employed, unless stated otherwise.

The nonlinear resistance R_{nl} considered here is associated with the flow separation from the tube's ends, both inside the actuator and at the actuator's outlet. The nonlinear acoustic impedance of orifices has been studied by, among others, Ingard and Labate,[15] Ingard and Ising,[16] and Thurston *et al.*[17] More recently Petculescu and Wilen[18] studied the nonlinear impedance for constrictions in a pipe. Thurston *et al.*[17] and Iguchi *et al.*[19] found that in many cases a quasi-steady approach for the nonlinear acoustic impedance can be

used. The instantaneous behavior of the nonlinear impedance for oscillatory flow is in that case equivalent to the minor loss for steady flow, which has an associated pressure drop related to the Bernoulli effect. In general, we can then write for the nonlinear resistance as function of time t (the impedance is now purely real):

$$R_{\mathrm{nl}}(t) = \frac{\rho K |U_{\mathrm{tube}}(t)|}{2S^2}, \quad K = K_{\mathrm{out}}, \quad U_{\mathrm{tube}}(t) > 0,$$

$$K = K_{\mathrm{in}}, \quad U_{\mathrm{tube}}(t) < 0, \tag{7}$$

with $U_{\mathrm{tube}}(t)$ the acoustic volume velocity in the tube, and K a constant called the loss parameter. The asymmetry of the geometry is accounted for by considering separate K_{in} and K_{out}. In Fig. 10 of Petculescu and Wilen[18] values for K_{in}, K_{out} of about 1 for short tubes, $L/d \approx 1$, to about 1.4 for long tubes, $L/d \approx 20$, are reported. Note here that these results apply for tube exits with a radius of curvature of 1.85 mm. Flow separation and hence the nonlinear loss parameter nevertheless depend somewhat on the radius of curvature, see also, e.g., Refs. 20 and 21. Especially when the radius of curvature becomes very small (less than a few tens of a millimeter), the nonlinear loss parameter can differ from the above mentioned values.

From analyzing the circuit of Fig. 2, the following two equations can be derived in the time-domain:

$$\frac{de(t)}{dt} = c_1 \frac{d^3U(t)}{dt^3} + c_2 \frac{d^2U(t)}{dt^2} + c_3 \frac{dU(t)}{dt} + c_4 U(t) - c_5 \frac{dU_{\mathrm{tube}}(t)}{dt} - c_6 U_{\mathrm{tube}}(t),$$

$$c_1 = \frac{L_e M_m}{S_d Bl},$$

$$c_2 = \frac{R_e M_m + L_e R_m}{S_d Bl},$$

$$c_3 = \frac{R_e R_m}{S_d Bl} + \frac{L_e S_d}{Bl C_{\mathrm{ac}}} + \frac{L_e}{S_d Bl C_m} + \frac{Bl}{S_d},$$

$$c_4 = \frac{R_e S_d}{Bl C_{\mathrm{ac}}} + \frac{R_e}{S_d Bl C_m},$$

$$c_5 = \frac{L_e S_d}{Bl C_{\mathrm{ac}}},$$

$$c_6 = \frac{R_e S_d}{Bl C_{\mathrm{ac}}}, \tag{8}$$

and

$$U(t) = C_{\mathrm{ac}} M_{\mathrm{ac,eff}} \frac{d^2U_{\mathrm{tube}}(t)}{dt^2} - cC_{\mathrm{ac}} M_{\mathrm{ac}} \,\mathrm{Im}(k) \frac{dU_{\mathrm{tube}}(t)}{dt} + \frac{\rho K C_{\mathrm{ac}}}{2S^2} |U_{\mathrm{tube}}(t)| \frac{dU_{\mathrm{tube}}(t)}{dt} + \left(\frac{\rho K C_{\mathrm{ac}}}{2S^2} \frac{d|U_{\mathrm{tube}}(t)|}{dt} + 1 \right) U_{\mathrm{tube}}(t). \tag{9}$$

In the latter equation, Eqs. (4) and (7) are used for substituting for the linear and nonlinear acoustic resistances.

Subsequently, the quantities $e(t)$, $U(t)$, and $U_{\mathrm{tube}}(t)$ are discretized to a number of N points. The system of equations (8) and (9) can now be solved numerically by a finite difference scheme, resulting in the following matrix equation:

$$\begin{bmatrix} \mathbf{U} \\ \mathbf{U}_{\mathrm{tube}} \end{bmatrix} = \mathbf{M}^{-1} \begin{bmatrix} \mathbf{D}_1 \mathbf{e} \\ \mathbf{0} \end{bmatrix},$$

$$\mathbf{M} = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 \\ \mathbf{m}_3 & \mathbf{m}_4 \end{bmatrix},$$

$$\mathbf{m}_1 = c_1 \mathbf{D}_3 + c_2 \mathbf{D}_2 + c_3 \mathbf{D}_1 + c_4 \mathbf{I}_{\mathbf{d}},$$

$$\mathbf{m}_2 = -c_5 \mathbf{D}_1 - c_6 \mathbf{I}_{\mathbf{d}},$$

$$\mathbf{m}_3 = -\mathbf{I}_{\mathbf{d}},$$

$$\mathbf{m}_4 = C_{\mathrm{ac}} M_{\mathrm{ac,eff}} \mathbf{D}_2 - cC_{\mathrm{ac}} M_{\mathrm{ac}} \,\mathrm{Im}(k) \mathbf{D}_1 + \frac{\rho C_{\mathrm{ac}}}{2S^2} [(K|\mathbf{U}_{\mathrm{tube}}|)_{\mathbf{d}} \mathbf{D}_1 + (\mathbf{D}_1 (K|\mathbf{U}_{\mathrm{tube}}|))_{\mathbf{d}}] + \mathbf{I}_{\mathbf{d}}. \tag{10}$$

Here, $\mathbf{e}$, $\mathbf{U}$, $\mathbf{U}_{\mathrm{tube}}$, and $|\mathbf{U}_{\mathrm{tube}}|$ are column vectors of length N representing the discretized version of the corresponding quantities. $\mathbf{D}_{1,2,3}$ are $N \times N$ matrices giving the first, second, and third order differential operators, respectively. Since an exact single period of a periodic (harmonic) voltage signal $e(t)$ will be considered as input, these operators satisfy a periodic condition. The differential operators employed in the present work are accurate to $O((\Delta t)^6)$, with Δt the discretization step in t. Furthermore, $\mathbf{I}_{\mathbf{d}}$ is the $N \times N$ identity matrix, and subscript $\mathbf{d}$ with a vector indicates an $N \times N$ matrix containing that vector on its diagonal. The term $K|\mathbf{U}_{\mathrm{tube}}|$ is evaluated at each discrete point as an asymmetrical K can be considered: $K_{\mathrm{in}} \neq K_{\mathrm{out}}$. Note that, in considering the linear resistance associated with wave damping, the wavenumber k corresponding to the imposed fundamental frequency is taken. This is a justifiable approximation, as it is found that the velocity amplitude at the higher harmonics is small compared to that at the fundamental frequency. The model, Eq. (10), is solved iteratively in a MATLAB implementation.

III. COMPARISON WITH EXPERIMENTAL RESULTS

In this section, model results for the momentum flux of the jet and, subsequently, the radiated sound are compared to experimental results. The asymmetric dipole synthetic jet actuator used consists of a custom loudspeaker (PSS 41 03978) connected to a square 6×6 cm^2 aluminum plate with a

TABLE I. Common parameters of the employed synthetic jet actuators.

Electrical resistance	R_e	5.70 Ω
Inductance	L_e	9.46×10^{-4} H
Force factor	Bl	5.36 N/A
Mechanical resistance	R_m	0.543 kg s^{-1}
Moving mass	M_m	4.70×10^{-3} kg
Compliance	C_m	5.42×10^{-4} m/N
Cone area	S_d	1.30×10^{-3} m^2
Front volume	V	1.20×10^{-5} m^3

thickness of 8 mm. A cavity is milled in the plate at the loudspeaker's side to accommodate the motion of the cone. At the other side a tube can be placed in a hole drilled in the middle of the plate. Different actuators are obtained by using tubes with different length and/or diameter. The common parameters of the actuators, i.e., that of the loudspeaker and the front volume V between speaker and tube are given in Table I.

A. Momentum flux

Neglecting viscous effects in the jet, the momentum flux I_{flow} through the plane perpendicular to the jet can be considered to be constant with distance from the outlet of the actuator. Since only the outward directed part of the fluid's acoustic motion at the tube exit contributes to the jet, the time-averaged momentum flux is

$$\overline{I_{\text{flow}}} = \frac{\rho}{S}\frac{1}{T}\int_{\substack{0,\\ U_{\text{tube}}(t)>0}}^{T} (U_{\text{tube}}(t))^2 dt. \tag{11}$$

For the numerical electroacoustic model, the time-averaged momentum flux is

$$\overline{I_{\text{flow}}} = \frac{\rho}{S}\frac{1}{N}\sum_{\substack{n=1,\\ \mathbf{U}_{\text{tube}}(n)>0}}^{N} (\mathbf{U}_{\text{tube}}(n))^2. \tag{12}$$

Experimentally, the time-averaged momentum flux can be found by letting the synthetic jet actuator blow on a balance, see Fig. 3. As the surface of the balance deflects the jet in the radial direction perpendicular to the original direction of the jet, the force exerted on the balance equals the momentum flux, provided that the jet is completely covered by the surface of the balance. Measurements and calculations of the momentum flux have been done for several configurations and different applied voltages. Moderate length brass tubes with diameter between 3 and 6 mm and length 30 and 60 mm as well as very long flexible plastic tubes with diameter 6 mm and length 1000 and 2030 mm have been used. The configurations are indicated by the diameter d and length L of the tube in millimeters as well as the rms value of the applied voltage. So, e.g., the configuration with $d=3$ mm, $L=60$ mm and rms voltage 3 V is denoted d3L60 3 V.

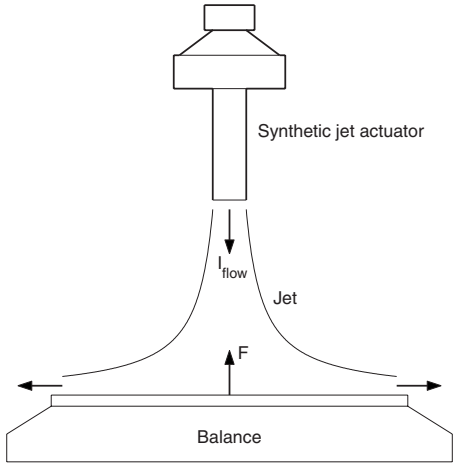


FIG. 3. Schematic drawing of set-up for measuring the momentum flux of the synthetic jet.

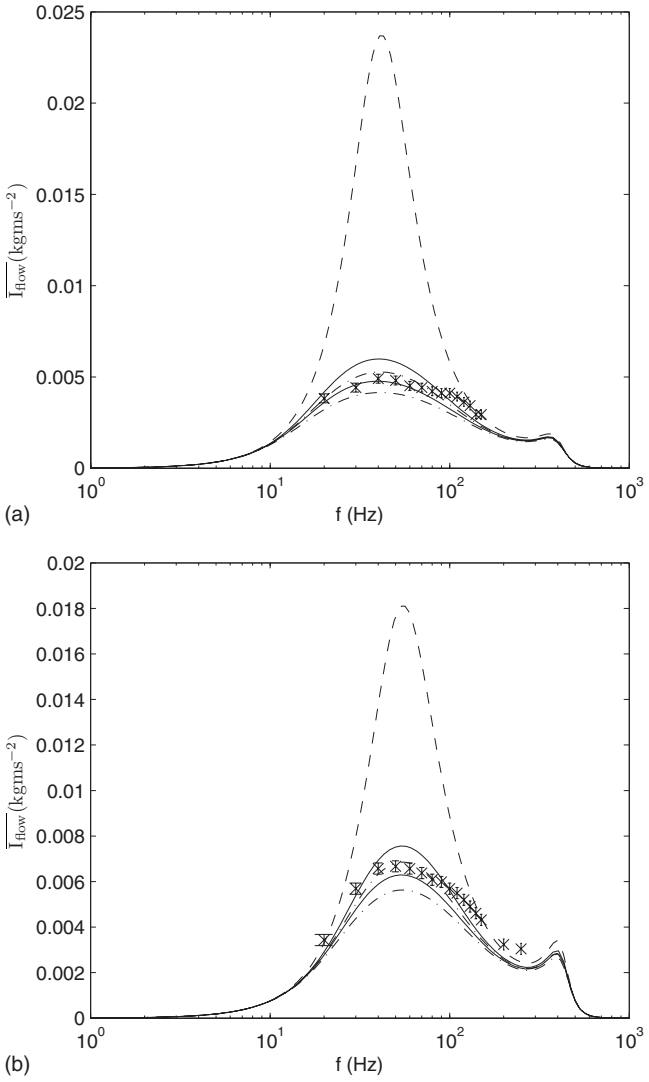


FIG. 4. Momentum flux for configurations: (a) d3L60 3 V and (b) d4L60 3 V. × markers: measurements; dashed lines: analytical model linear; dashed-dotted lines: analytical model nonlinear; solid lines: numerical model nonlinear. Higher and lower values are for $K_{\text{in}}=K_{\text{out}}=1$ and $K_{\text{in}}=K_{\text{out}}=1.4$, respectively.

1. Moderate length tubes

Results for the configurations with moderate length brass tubes, namely, d3L60 and d4L60 at 3 V, and d6L30 and d6L60 at 2 V, are shown in Figs. 4 and 5.

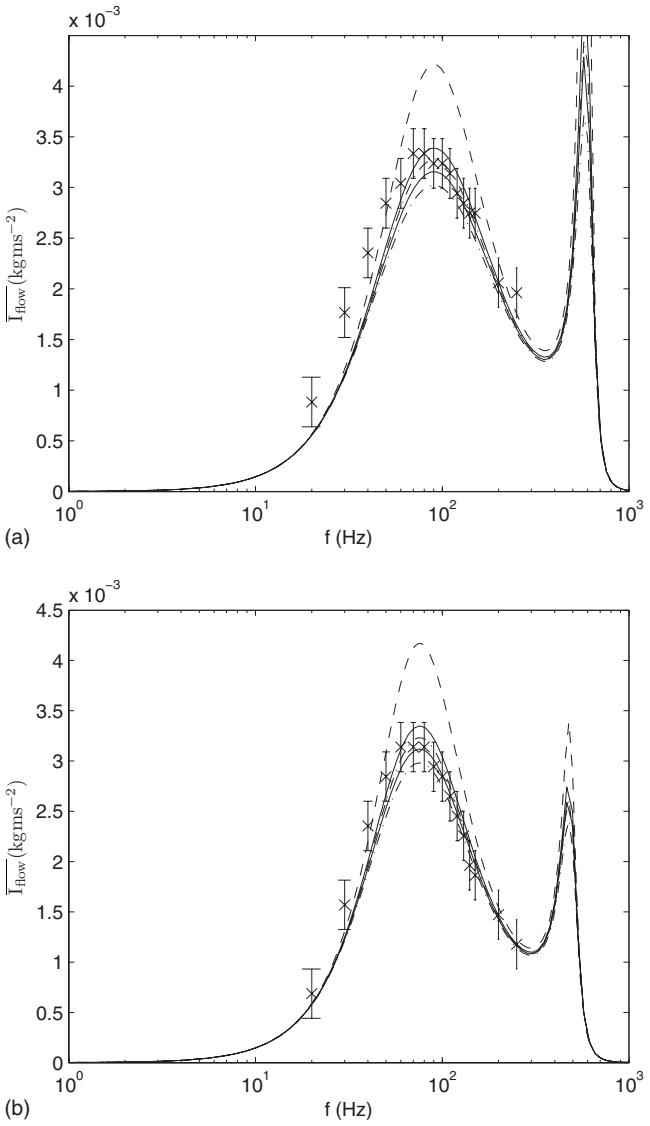


FIG. 5. As in Fig. 4 for configurations: (a) d6L30 2 V and (b) d6L60 2 V.

Both calculations with the current numerical model of Sec. II, including nonlinear losses, and an analytical lumped element electroacoustic model (in frequency domain) without and with nonlinear losses are done. In the latter case the nonlinear flow resistance is set constant to $R_{nl}=\rho K|\hat{U}_{tube}|/(2S^2)$, with $\hat{U}_{tube}$ the complex amplitude of the volume velocity in the tube (see also Ref. 11). The number of points used for the numerical calculations is $N=100$, using more points gave no significant difference in results. It was verified that without accounting for nonlinear losses the results obtained from the numerical model are identical to those obtained by the analytical model. In all cases wave damping is accounted for by means of Eq. (5). Results from both the analytical and the numerical model are obtained with the nonlinear loss parameter $K_{in}=K_{out}=1$ (giving the corresponding higher values for the momentum flux) and $K_{in}=K_{out}=1.4$ (giving the lower values for the momentum flux). These values for K_{in}, K_{out} correspond to, respectively, short tubes, $L/d\approx 1$, and long tubes, $L/d\approx 20$, see Ref. 18. The iteration process in the analytical model was stopped once the absolute value of the difference in $\hat{U}_{tube}$ between two successive iterations divided by the absolute value of $\hat{U}_{tube}$ was less than 10^{-6}. For the numerical model iteration was stopped once at each discrete point, the absolute value of the difference in $\mathbf{U}_{tube}$ between two successive iterations divided by the maximum absolute value of $\mathbf{U}_{tube}$ was less than 10^{-6}. For both models the number of iterations needed for the present calculations was about 30 at maximum.

Generally, a good prediction of the momentum flux is obtained when including the nonlinear resistance. The smaller the diameter of the tube, the higher the flow velocity, and the higher the (effect of) nonlinear resistance. This is especially apparent for the d3L60 configuration, where the momentum flux differs by about a factor 5, and the velocity thus by a factor ~2.2, between the calculations with and without nonlinear losses. Furthermore, it was found that the effect of damping in wave propagation is much smaller than that of the nonlinear losses. In the analytical model, the nonlinear resistance is considered as constant, and the amplitude of the velocity is used to calculate it. So, in fact, the maximum value of the nonlinear resistance is used over a complete cycle, yielding an overestimation. This results in the lower value of the momentum flux calculated by the analytical model compared to that calculated by the numerical method, and consequently less good agreement with experimental results. Note here that of course the nonlinear resistance for the analytical model could also be calculated from the rms value of the velocity. However, it was found that this, in fact, in turn gives an underestimation of the nonlinear loss (the results for $K_{in}=K_{out}=1.4$ would in that case be about equal to the results for $K_{in}=K_{out}=1$ as displayed in Figs. 4 and 5). Furthermore, when considering the relative length L/d of the tubes, there is a good correspondence between the measured momentum flux and the one calculated with the numerical method using either $K_{in}=K_{out}=1$ for the shorter tubes or $K_{in}=K_{out}=1.4$ for the longer tubes. Wave damping is included by means of Eq. (5). As the effect is relatively small for these short tubes using Eq. (6) instead gave no significant difference.

2. Very long tubes

For the configurations with a very long tube, caution has to be taken in using the low-frequency approximation, as $(\omega L/c)^2 \ll 1$ is actually not valid everywhere: for the 1000 and 2030 mm tubes $\omega L/c=1$ at a frequency of $f=55$ Hz and $f=27$ Hz, respectively. Two different linear calculations with the analytical model have been carried out. In the first the tube is considered as a transmission line; in the second it is considered as an acoustical mass, i.e., low-frequency approximation. Calculations including nonlinear losses are carried out with the analytical as well as the numerical model, and are both in low-frequency approximation. The nonlinear loss parameter is taken $K_{in}=K_{out}=1.4$. Since the tubes are very long, wave damping will be much more important. The model calculations and the experimental results are given in Fig. 6. Comparing results from the linear model with and without low-frequency approximation, it is observed that with the low-frequency approximation still the first resonance is predicted quite well. When accounting for the non-

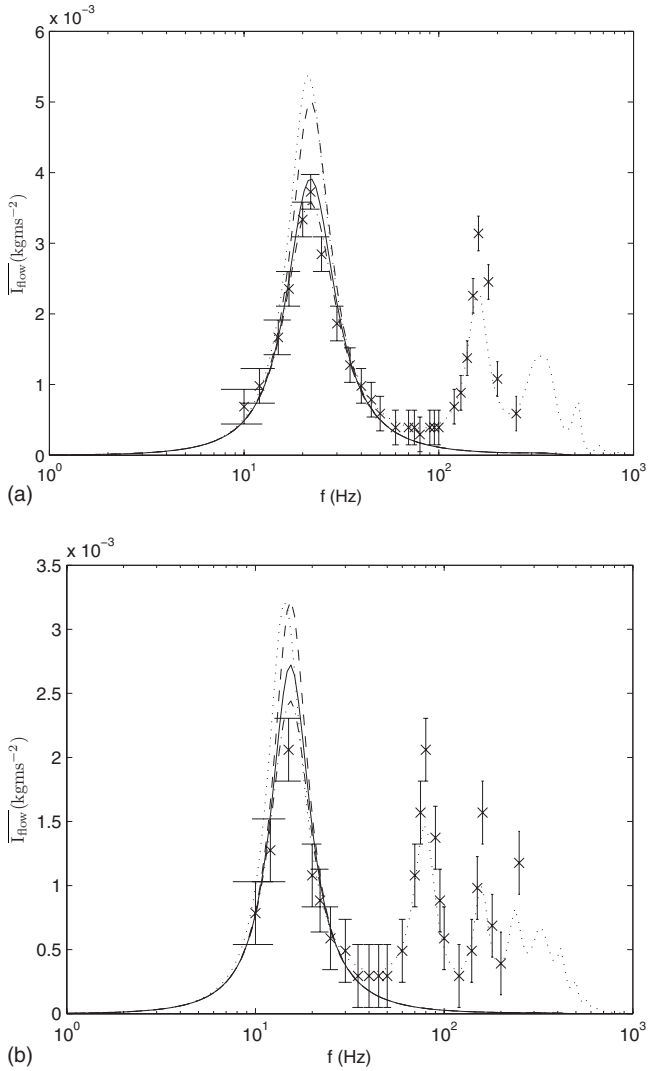


FIG. 6. Momentum flux for configurations: (a) d6L1000 3 V and (b) d6L2030 3 V. × markers: measurements; dotted lines: analytical model, linear; dashed lines: analytical model, linear, low frequency approximation; dashed-dotted lines: analytical model, nonlinear, low frequency; solid lines: numerical model, nonlinear, low frequency. Nonlinear loss parameter of $K_{in}=K_{out}=1.4$ is used.

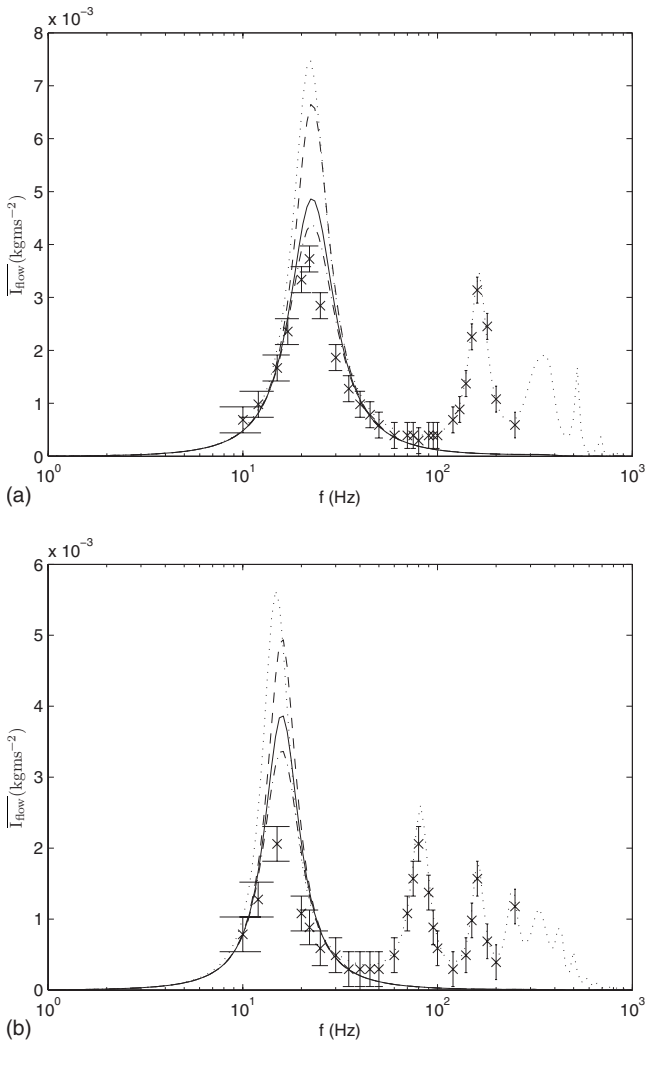


FIG. 7. Momentum flux for configurations: (a) d6L1000 3 V and (b) d6L2030 3 V. As in Fig. 6, but with wave damping according to Eq. (6) instead of Eq. (5).

linear losses, an improved prediction of the measured momentum flux at the first resonance is obtained. However, for the resonances at higher frequencies it is seen that the momentum flux is already underestimated by the linear model when using Eq. (5). It was found that using the wave damping model according to Eq. (6), valid for high shear numbers, yields better results for the higher frequencies, but worse results at the first resonance. This is depicted in Fig. 7. The shear number ranges here from ~6 for $f=10$ Hz to ~30 for $f=250$ Hz.

B. Radiated sound

The nonlinear flow resistance results in the generation of higher harmonics of the fundamental frequency with which the loudspeaker is driven, i.e., distortion occurs. In this section, model results for the sound power radiated by the asymmetric dipole cooler will be compared to experimental results.

With the current model, the volume velocity at the loudspeaker, $U(t)$, and at the tube outlet, $U_{tube}(t)$, are calculated in the time domain. These volume velocities can be considered in the frequency domain using a Fourier transform. The complex amplitudes–as function of angular frequency ω—of the volume velocity at the back of the loudspeaker and at the tube outlet are $-\hat{U}(\omega)$ and $\hat{U}_{tube}(\omega)$, respectively.

Considering the tube outlet and loudspeaker back as monopole radiators, it can be derived that the time-averaged total radiated acoustic power as function of frequency equals

$$\bar{P}(\omega)=\frac{k_0^2\rho c|\hat{U}_m(\omega)|^2}{8\pi}+\frac{k_0^4\rho c|p_d(\omega)|^2}{24\pi}, \tag{13}$$

see, e.g., Ref. 22. This corresponds to the radiated power of a monopole with strength (i.e., volume velocity):

$$\hat{U}_m(\omega)=\hat{U}_{tube}(\omega)-\hat{U}(\omega), \tag{14}$$

plus that of a dipole, with dipole strength:

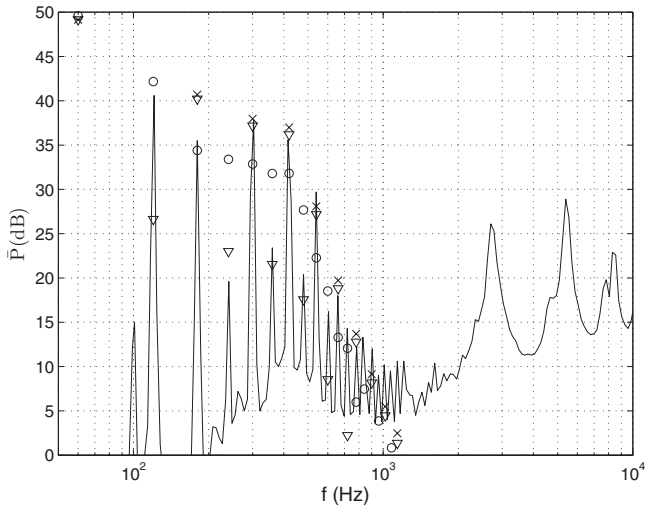


FIG. 8. Power spectrum of radiated sound for the d4L60 driven at 60 Hz and 3 V rms voltage. Solid line: experimental result. Model results are for × markers: symmetric nonlinear resistance with $K_{in}=K_{out}=1.2$. ○ markers: fully asymmetric nonlinear resistance with $K_{in}=0$, $K_{out}=1.2$. ▽ markers: asymmetric nonlinear resistance with $K_{in}=0.9$, $K_{out}=1.2$.

$$p_d(\omega) = (\hat{U}(\omega) + \hat{U}_{\text{tube}}(\omega))\frac{a}{2}. \tag{15}$$

Here, a is the distance between the loudspeaker back and the tube outlet. The above holds in the far-field approximation, i.e., $k_0 r$ large, with r the distance to the actuator, and for $k_0 a$ small.

Power spectra of the radiated sound for various configurations have been measured in a 227 m^3 reverberant room. Frequencies below 100 Hz are filtered out, as below this frequency the sound field is not diffuse enough to give accurate results. The spectrum for the configuration d4L60 driven at 60 Hz and 3 V rms voltage is shown in Fig. 8. The acoustic power is expressed in decibels relative to a reference of 10^{-12} W. The higher harmonics at 120 Hz, 180 Hz,…, etc., can be clearly discerned. Also the tube resonances, giving the broad peaks at ~2800 Hz, ~5500 Hz,…, etc., are observed. Additionally, the figure depicts some results obtained from the numerical model. Here, calculations have been done for a symmetric nonlinear resistance with nonlinear loss parameters $K_{in}=K_{out}=1.2$, as well as for two cases with asymmetric nonlinear resistance. For the symmetric nonlinear resistance, only odd harmonics are generated. However, from the experimental results it can be clearly seen that, although the odd harmonics are dominant in level, even harmonics occur as well. This suggests that there is some asymmetry in the nonlinear resistance. The effect of this asymmetry is investigated in Fig. 8 by considering a "full" asymmetrical configuration with nonlinear loss parameters $K_{in}=0$, $K_{out}=1.2$ (thus only nonlinearity at outflow), and an intermediate configuration with $K_{in}=0.9$, $K_{out}=1.2$. For both cases even as well as odd harmonics are indeed observed. Comparison with the experimental results shows here that the full asymmetrical case gives a clear overestimate of the even harmonics, whereas the intermediate case gives a better prediction for the higher harmonics (except for the first higher harmonic, which is now underestimated). Note here that the full asymmetrical case also gives a deviant prediction of the momentum flux: $\overline{I_{\text{flow}}}=9.3\times10^{-3}$ kg m s^{-2}, whereas experimentally $\overline{I_{\text{flow}}}=6.6\times10^{-3}$ kg m s^{-2}, and for the intermediate asymmetrical case: $\overline{I_{\text{flow}}}=7.2\times10^{-3}$ kg m s^{-2}. So, a (very plausible) slight asymmetry in the nonlinear resistance yields the generation of both even and odd harmonics, and gives a reasonable quantitative agreement with experimental results. This also means that in principle the values of the nonlinear loss parameters K_{in}, K_{out} would have to be known quite accurately to obtain an accurate quantitative prediction of the complete (power) spectrum of the radiated sound. Furthermore, regarding the volume velocity, the higher harmonics are very small compared to the fundamental frequency (at least one order in magnitude smaller). Therefore, other nonlinearities in the system, e.g., nonlinearities in the suspension of the loudspeaker, that have a negligible effect on the total volume velocity—and thus on the blowing velocity of the actuator—may still be important as well for the radiation of higher harmonics. In principal, these loudspeaker nonlinearities cannot be isolated in the measurements on the synthetic jet actuator. Also, the enclosure, i.e., the volume and tube of the actuator, has influence on the nonlinearities of the loudspeaker. Nevertheless, in order to obtain some indication, distortion spectra of the loudspeaker without any enclosure (dipole configuration) as well as the loudspeaker in a closed box of 0.9 l (monopole configuration) have been measured at a driving frequency of 60 Hz. As loudspeaker nonlinearities are greatly dependent on the cone excursion, the applied voltage for both cases was chosen such that the excursion is the same as for the synthetic jet actuator configuration. Determined from lumped element modeling, this yielded rms voltages of 2.0 and 1.9 V, respectively, to have a loudspeaker cone excursion amplitude of 1.03 mm. For the first three higher harmonics, thus up to 240 Hz, the observed level, relative to the fundamental, in these spectra was roughly of the same order as in the spectrum of the synthetic jet actuator, see Fig. 8. For the other higher harmonics, the level was generally at least 10 dB lower than in the case of the synthetic jet actuator. From this it can tentatively be concluded that in the spectrum of Fig. 8 the distortion components above the third higher harmonic at 240 Hz are mainly generated by the nonlinear flow resistance, whereas for the first three higher harmonics loudspeaker nonlinearities play a significant role as well.

Finally, since both in the experiment and in the model the odd harmonics are dominant in level, and globally a reasonable agreement is obtained for these harmonics, it is found that a fair estimate for the total power of the radiated higher harmonics can be obtained for most cases. This is shown in Table II. Here, in determining the total measured sound power, the contribution of the tube resonances is omitted. For the calculated sound power nonlinear loss parameter $K_{in}=K_{out}=1.2$ is taken for all cases. In Table II the measured total radiated power by the harmonics is usually larger than the calculated one. This is partly due to the fact that the first higher harmonic is underestimated by the model (for a symmetrical nonlinear flow resistance, it is absent in model results).

TABLE II. Total power radiated by high harmonics for various configurations, measured in a reverberant room and calculated from the numerical model with $K_{in}=K_{out}=1.2$.

Configuration	Total power measurement (dB)	Total power model (dB)
d4L30 60 Hz 3 V	50.4	48.2
d4L60 60 Hz 3 V	44.6	43.8
d6L30 60 Hz 3 V	45.7	42.6
d6L30 70 Hz 3 V	48.9	46.0
d6L30 80 Hz 3 V	46.9	49.4
d6L60 60 Hz 3 V	47.7	40.8
d6L60 70 Hz 3 V	49.2	44.3
d6L60 80 Hz 3 V	47.3	46.3

IV. CONCLUSIONS

An electroacoustic model, in which nonlinear flow resistance associated with flow separation can be included, is presented for an asymmetric dipole synthetic jet actuator. The numerical model employs a finite difference scheme in the time domain, allowing for a full inclusion of the nonlinear resistance.

Model calculations for the time-averaged momentum flux of the jet are compared to experimental results for various synthetic jet actuator configurations. It is shown that a significantly improved prediction of the momentum flux—and thus flow velocity—of the jet is obtained when including the nonlinear resistance. Here, the current model performs slightly better than a more common analytical frequency domain electroacoustic model, in which the nonlinear resistance can only be considered as a constant.

As opposed to such an analytical model, with the numerical model also higher harmonics, i.e., distortion components, generated as a result of the nonlinear resistance, can be calculated. For a single configuration the power spectrum of radiated sound, measured in a reverberant room, is compared to the calculated one. A reasonable agreement is obtained when assuming a plausible slight asymmetry in the nonlinear resistance (i.e., a difference between inflow and outflow). It was also found that probably distortion above the third higher harmonic is mainly caused by the nonlinear flow resistance, whereas for the first three higher harmonics loudspeaker nonlinearities play a significant role as well. Furthermore, in principle, for an accurate prediction of the distortion the exact nonlinear resistance at in- and outflow would have to be known. Nevertheless, a fair prediction of the total radiated power of higher harmonics is obtained, which is less sensitive to the exact value of the nonlinear resistance.

The presented method can also in general be used for loudspeaker systems containing a bass port.

ACKNOWLEDGMENT

The authors would like to thank Celine Nicole, Clemens Lasance, Joris Nieuwendijk and Ronald Aarts for their collaboration.

[1] A. Glezer and M. Amitay, "Synthetic jets," Annu. Rev. Fluid Mech. **34**, 503–529 (2002).
[2] K. Mosheni, "Pulsatile vortex generators for low-speed maneuvering of small underwater vehicles," Ocean Eng. **33**, 2209–2223 (2006).
[3] Z.-b. Luo and Z.-x. Xia, "A novel valve-less synthetic-jet-based micropump," Sens. Actuators, A **122**, 131–140 (2005).
[4] R. Mahalingam, S. Heffington, L. Jones, and R. Williams, "Synthetic jets for forced air cooling," Electronics Cooling, May 2007.
[5] C. J. M. Lasance and R. M. Aarts, "Synthetic jet cooling Part I: Overview of heat transfer and acoustics," 24th IEEE Semi-Therm Symposium (2008), pp. 20–25.
[6] J. Punch, "Point-of-source thermal management," IEEE Sixth EuroSimE Conference (2005), pp. 688–694.
[7] R. M. Aarts, J. A. M. Nieuwendijk, and O. Ouweltjes, "Efficient resonant loudspeakers with large form-factor design freedom," J. Audio Eng. Soc. **54**, 940–953 (2006).
[8] R. Holman, Y. Utturkar, R. Mittal, B. L. Smith, and L. Cattafesta, "Formation criterion for synthetic jets," AIAA J. **43**, 2110–2116 (2005).
[9] R. M. Aarts, J. A. M. Nieuwendijk, and A. J. J. Wismans, "Pulsating cooling system," Patent Application No. WO2008065602 (2008).
[10] R. M. Aarts and C. J. M. Lasance, "Cooling device and electronic device comprising such a cooling device," Patent Application No. WO2007107921 (2007).
[11] Q. Gallas, R. Holman, T. Nishidaa, B. Carroll, M. Sheplak, and L. Cattafesta, "Lumped element modeling of piezoelectric-driven synthetic jet actuators," AIAA J. **41**, 240–247 (2003).
[12] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustic*, 3rd ed. (Wiley, New York, 1982).
[13] H. Tijdeman, "On the propagation of sound in cylindrical tubes," J. Sound Vib. **39**, 1–33 (1975).
[14] C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949).
[15] U. Ingard and S. Labate, "Acoustic circulation effects and the nonlinear impedance of orifices," J. Acoust. Soc. Am. **22**, 211–218 (1950).
[16] U. Ingard and H. Ising, "Acoustic nonlinearity of an orifice," J. Acoust. Soc. Am. **42**, 6–17 (1967).
[17] G. B. Thurston and C. E. Martin, "Periodic fluid flow through circular orifices," J. Acoust. Soc. Am. **25**, 26–31 (1953).
[18] A. Petculescu and L. A. Wilen, "Oscillatory flow in jet pumps: Nonlinear effects and minor losses," J. Acoust. Soc. Am. **113**, 1282–1292 (2003).
[19] M. Iguchi, M. Ohmi, and K. Maegawa, "Analysis of free oscillating flow in a U-shaped tube," Bull. JSME **25**, 165–172 (1982).
[20] J. H. M. Disselhorst and L. van Wijngaarden, "Flow in the exit of open pipes during acoustic resonance," J. Fluid Mech. **99**, 293–319 (1980).
[21] M. Atig, J.-P. Dalmont, and J. Gilbert, "Saturation mechanism in clarinet-like instruments; The effect of the localised nonlinear losses," Appl. Acoust. **65**, 1133–1154 (2004).
[22] L. L. Beranek, *Acoustics* (Acoustical Society of America, Melville, 1996).

The interference component of the acoustic field corresponding to the Long-Range Ocean Acoustic Propagation Experiment

Natalie S. Grigorieva and Gregory M. Fridman
St. Petersburg State Marine Technical University, 3 Lotsmanskaya Strasse, 190008 St. Petersburg, Russia

James A. Mercer, Rex K. Andrew,[a] and Michael A. Wolfson
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

Bruce M. Howe
University of Hawaii, Honolulu, Hawaii 96822

John A. Colosi
Naval Postgraduate School, Monterey, California 93943

(Received 12 June 2008; revised 22 January 2009; accepted 23 January 2009)

Propagation of energy along the sound channel axis cannot be formally described in terms of geometrical acoustics due to repeated cusped caustics along the axis. In neighborhoods of these cusped caustics, a very complicated interference pattern is observed. Neighborhoods of interference grow with range and overlap at long ranges. This results in the formation of a complex interference wave—the axial wave—that propagates along the sound channel axis like a wave belonging to a crescendo of near-axial arrivals. The principal properties of this wave are calculated for the actual space-time configuration realized during a 2004 long-range propagation experiment conducted in the North Pacific. The experiment used *M*-sequences at 68.2 and 75 Hz, transmitter depths from 350 to 800 m, and ranges from 50 to 3200 km. Calculations show that the axial wave would be detectable for an optimal geometry—both transmitter and receiver at the sound channel axis—for a "smooth" range-dependent sound speed field. The addition of sound speed perturbations—induced here by simulated internal waves—randomizes the acoustic field to the extent that the axial wave becomes undetectable. These results should be typical for mid-latitude oceans with similar curvatures about the sound speed minimum.
 [DOI: 10.1121/1.3082112]

PACS number(s): 43.30.Dr, 43.30.Qd [RCG] Pages: 1919–1929

I. INTRODUCTION

The world's mid-latitude oceans possess a well-known deep sound channel where decreasing sound speed—due to decreasing temperature—with increasing depth gives way to increasing sound speed—due to increasing hydrostatic pressure—with increasing depth. This sound channel has long been exploited in studies of long-range sound ocean acoustic propagation. Often, the most prominent characteristics of sound propagation in this channel are early arrivals well described by geometrical acoustic theory, followed by a crescendo of energy that propagates along the center, or axis, of the channel. It is impossible to explain this late-arriving energy using geometrical acoustics because of the presence of repeated cusped caustics along the axis. Figure 1 illustrates a typical representation of caustics for these kinds of oceans.

In the neighborhood of an isolated cusp, the acoustic field is described by the Pearcey integral.[1] The extent of the cusp neighborhood where the geometrical acoustic formulas are not applicable increases with range,[2] and at a certain propagation range, the neighborhoods of adjacent cusps overlap. This results in a more complicated diffractive pattern that cannot be described by the Pearcey integral. The interference of the wave fields that correspond to near-axial rays leads to formation of a coherent structure that propagates along the axis like a wave. This component of the acoustic field is called the "axial wave." The range at which the cusp neighborhoods begin to overlap depends on frequency and sound speed profile properties, such as the curvature of the profile at the axis depth. When the distance between the transmitter and receiver increases, the number of caustic cusps participating in the axial wave formation increases as well.

The term axial wave appeared in the early 1970s in papers by Buldyrev[3] and Buldyrev *et al.*[4] Those studies were done for the simplest medium model with the purpose of obtaining a global description of propagation in the presence of cusped caustics.

For a range-independent ocean model where the variables are separable, the formula for the axial wave is derived by transforming the contour integral that gives the exact solution of the point transmitter problem.[5] Analysis of the resulting integral representation of the axial wave shows that it includes the solutions of the Helmholtz equation localized near the sound channel axis and a weight function that does not depend on medium properties. The main advantage of the

[a]Electronic mail: rex@apl.washington.edu

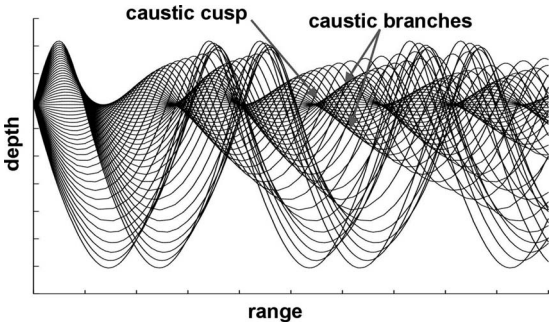


FIG. 1. Caustic features at relatively short ranges $0 \leqslant x \leqslant 150$ km, as illustrated by ray tracing. The sound speed profile (assumed in this example to be independent of range) is typical for the deep ocean at mid-latitudes; the transmitter (at zero range) is located at the sound channel axis (1 km), where the sound speed reaches its minimum.

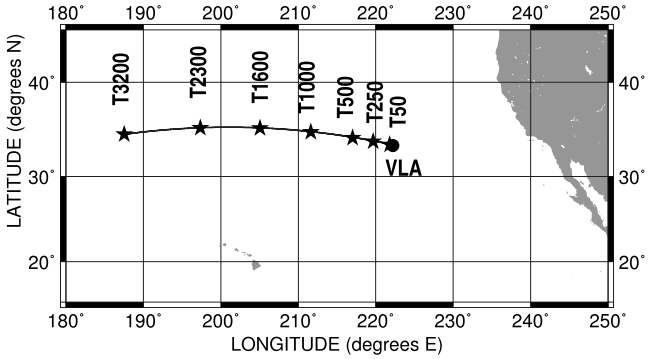


FIG. 2. LOAPEX geometry. The solid line (a geodesic arc on an oblate spheroidal model earth) shows the main LOAPEX path where stars indicate the seven stations: T50, T250, T500, T1000, T1600, T2300, and T3200. The VLA is marked by a circle.

derived formula for the axial wave is the possibility to generalize it to a range-dependent ocean, as was done by Grigorieva and Fridman.[6]

In the time-of-arrival pattern (the acoustic field as a function of depth and time, at a given range), the axial wave formed by rays with small launch angles is a component of the late-arriving energy. In long-range ocean acoustic propagation, the axial crescendo accounts for a significant portion of the radiated energy but is not currently utilized in acoustical oceanography problems. It is therefore important and useful to have a mathematical model for this component of the near-axial finale.

In this paper, the principal features of the axial wave are calculated for several increasingly realistic ocean medium models. These ocean models are derived from actual conditions encountered during the Long-Range Ocean Acoustic Propagation Experiment (LOAPEX), conducted in the North Pacific from 10 September to 10 October 2004. The goal of this paper is to demonstrate these calculations and to determine whether or not the axial wave could have been detected given optimal transmitter/receiver geometries and actual LOAPEX transmitter/receiver geometries. The general outline of this paper is as follows. In Sec. II we describe the LOAPEX sound speed data as well as the calculation of a smooth two-dimensional sound speed field approximation. In Sec. III we discuss the dependence of the main properties of the axial wave on propagation range, depth of the transmitter relative to the depth of the sound channel axis, and frequency for two ocean models: first, a smooth sound speed approximation and, second, the smooth approximation with additive perturbations representative of internal-wave induced variability. In Sec. IV we summarize the principal results. Appendix A includes the integral representations of the axial wave for range-independent and range-dependent ocean models. In Appendix B we give a short description of the acoustic fluctuations due to internal waves that are used for simulations in this paper.

II. SMOOTH TWO-DIMENSIONAL MODEL OF LOAPEX SOUND SPEED FIELD

A detailed summary of LOAPEX is given by Mercer *et al.*[7] The experiment geometry is shown in Fig. 2. Seven transmit stations, T50, T250, T500, T1000, T1600, T2300, and T3200, were located at ranges of 50, 250, 490, 990, 1600, 2300, and 3200 km, respectively, from a vertical line array (VLA). At each of these seven stations, the LOAPEX acoustic transmitter was suspended from the ship to conduct periodic transmissions for several hours. Multiple transmitter depths were used at each of these seven stations, involving 350, 500, and/or 800 m. At the transmitter depth of 800 m, the transmitted signal had a carrier frequency of 75 Hz, and at transmitter depths of 350 and 500 m, it had a carrier frequency of 68.2 Hz. The interelement spacing on the 20-element VLA was 35 m throughout the central region of the sound channel axis.

Figure 3 shows the upper ocean sound speed profiles obtained from the conductivity-temperature-depth (CTD) measurements at six stations (T250, T500, T1000, T1600, T2300, and T3200). The formula for the axial wave (see Appendix A) requires derivatives to fourth order of the sound speed $c(x,z)$ with respect to both range and depth. These were obtained by modeling the actual measurement profiles with a smooth representation, constructed as follows.

A 14th-order polynomial was fit in the least squares sense to each sound speed profile over the depth interval $[350,1500]$ m. This order was chosen because it gave reasonable fidelity to the actual measurements without excessive complexity. Figure 4 shows both the sound speed profile at station T50 and the corresponding polynomial approximation. The sound speed profile at the VLA was represented by that at station T50.

Using the polynomial sound speed profiles from these seven stations, the "smooth" two-dimensional sound speed field was calculated by generating additional sound speed profiles every 50 km in range using linear interpolation and then by interpolating the total sound speed profile data set (polynomials plus interpolated) using two-dimensional fifth-order splines. The calculated sound channel axis depth is shown in Fig. 5, where dots correspond to the depth of the transmitter at its closest to the channel axis. At stations T3200 and T2300, this depth was 500 m, at station T1600 this depth was 350 m, and at stations T1000, T500, T250, and T50 this depth was 800 m.

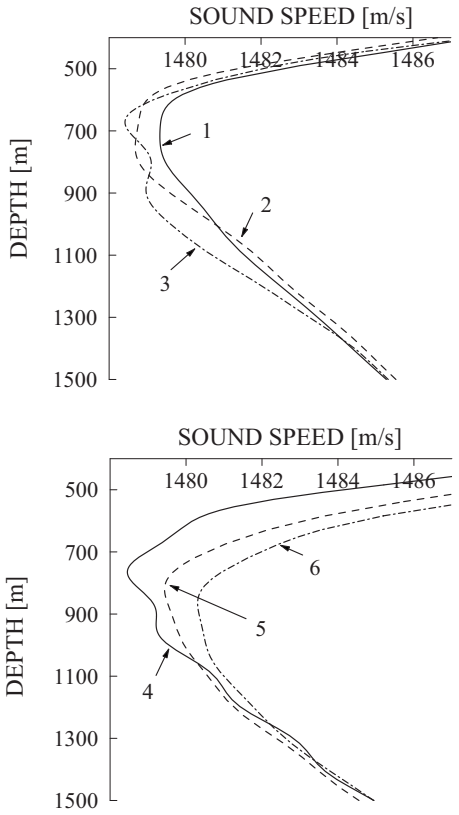


FIG. 3. Upper ocean sound speed profiles obtained from CTD data at six stations. The upper panel contains the sound speed profiles for stations T250 (1), T500 (2), and T1000 (3), and the lower panel contains the sound speed profiles for stations T1600 (4), T2300 (5), and T3200 (6).

The VLA hydrophone nearest the sound speed minimum (635.678 m) had a median depth over the entire deployment of 621.72 m, a location roughly 14 m from the channel axis.

III. MAIN PROPERTIES OF THE AXIAL WAVE

The integral representations of the axial wave for range-independent and range-dependent ocean models are written out in Appendix A.

A. Dependence on propagation range

Range dependence for a range-independent ocean has previously been examined[8] for the Munk canonical sound

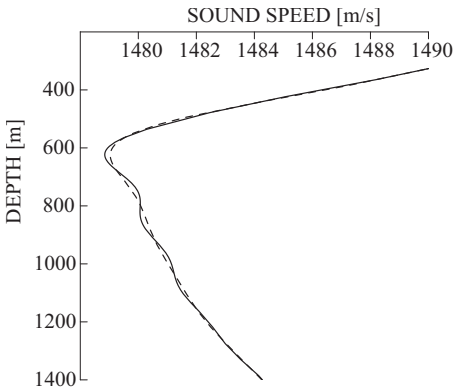


FIG. 4. Sound-speed profile obtained from CTD measurement at station T50 (solid) and its 14th degree polynomial approximation (dashed).

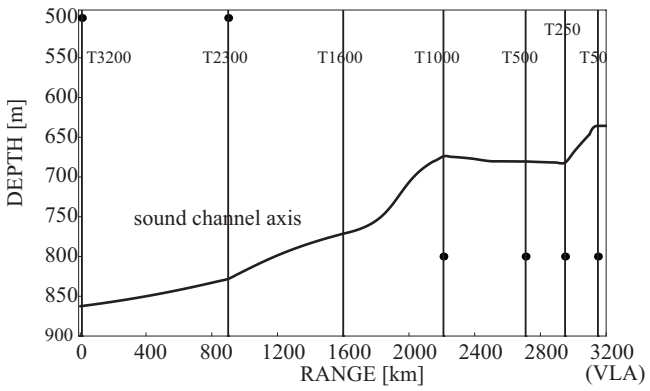


FIG. 5. The sound channel axis for the smooth two-dimensional sound speed field approximation. Dots show the depth of the transmitter when closest to the channel axis. No reference marks are shown for station T1600 because the only transmitter deployment depth, 350 m, was too far from the channel axis to generate a significant axial wave (see Sec. III B).

speed profile[9] and the average sound speed profile from the Acoustic Engineering Test (AET). The effect of the range-dependent LOAPEX sound speed profile is characterized by the transmission loss (TL) in Fig. 6. For this calculation, the transmitter is placed at the station T3200 sound channel axis depth, and the frequency is 68.2 Hz. The receiver is assumed to be located at the sound channel axis in the 50 km range interval adjoining the VLA. The TL varies from 117 to 120 dB. The dashed line in Fig. 6 indicates a range dependence of $x^{-3/4}$ for the wave amplitude. Thus, the axial wave decreases with range faster than the low-order modes. Discontinuities in the TL take place at distances $x=3154.58$, 3163.74, 3172.92, 3182.26, and 3191.31 km, where the number of waves corresponding to rays with small launch angles participating in the axial wave formation changes by 1. The transmitter level in LOAPEX (Ref. 7) was about 194 dB, and the ambient noise level throughout the sound channel at the VLA was roughly 70–80 dB re 1 μPa^2/Hz at 75 Hz. Based on this calculation (which neglects internal waves, see below) the axial wave would have been detectable if the transmitter and receiver had been located at the sound channel

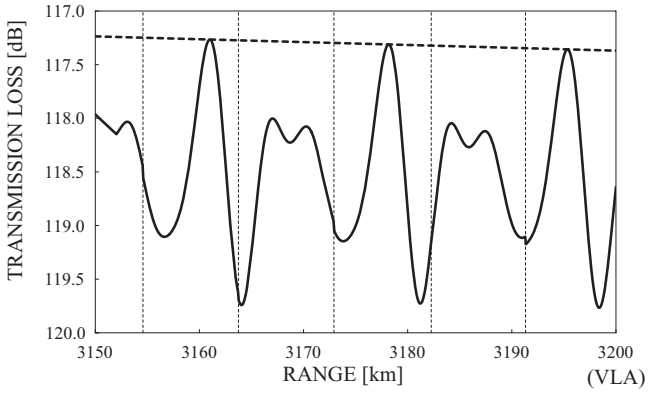


FIG. 6. Axial wave TL at the channel axis over the final 50 km of the LOAPEX path for a signal launched from a transmitter located at the station T3200 channel axis (862.535 m). The range-dependent sound speed is represented by a smoothed approximation of actual sound speed measurements. The frequency is 68.2 Hz. The heavy dash line represents an $x^{-3/4}$ dependence in amplitude. The vertical dashed lines indicate the locations of discontinuities in the TL (see text).

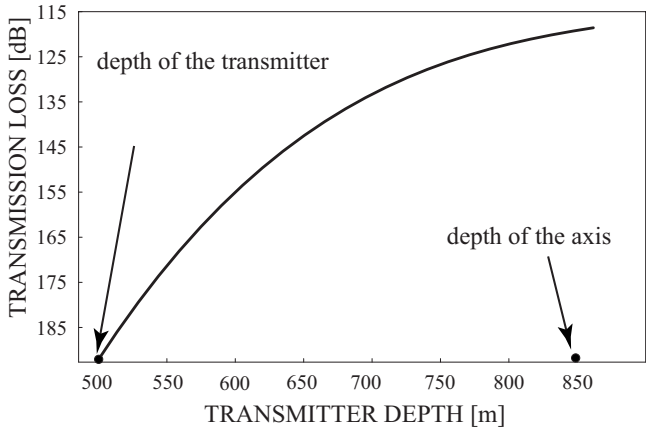


FIG. 7. Dependence of the axial wave TL on the depth of the transmitter placed at 3200 km from the receiver. (The sound speed minimum is at 862.535 m at station T3200.) The frequency is 68.2 Hz. The receiver is at the depth of the sound speed minimum (635.678 m) at the location of the VLA.

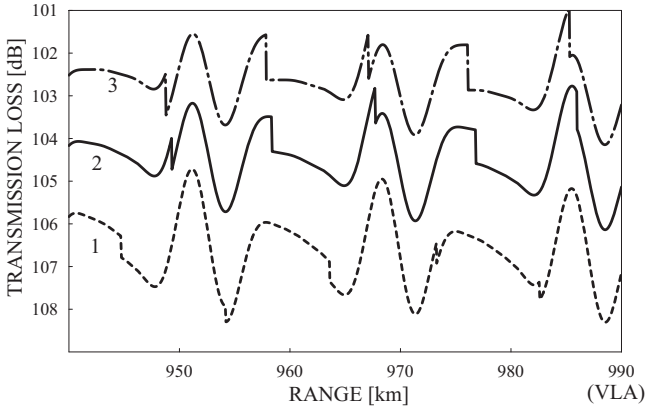


FIG. 8. TL for an axial wave launched from a transmitter at the sound channel axis (674.026 m) at station T1000. TL is predicted at the channel axis over the final 50 km. Curve (1) corresponds to 56.25 Hz, (2) to 75 Hz, and (3) to 93.75 Hz.

axis, even at a range of 3200 km. (This assumes some form of pulse compression post-processing gain, as is commonly used in long-range propagation experiments.[9])

B. Dependence on transmitter depth

Depth dependence at different propagation ranges for a range-independent ocean has previously been examined[8] for the Munk canonical sound speed profile[9] and the average sound speed profile from AET. The effect of depth dependence for the range-dependent LOAPEX sound speed profile is characterized by the TL in Fig. 7. For this calculation, the transmitter is at station T3200 and the depth is varied from the sound channel axis (862.535 m) up to 500 m, the depth of the actual transmitter deployment. The frequency is 68.2 Hz. The receiver is located on the sound channel axis. The TL increases rapidly when the distance between the transmitter and the channel axis increases. At the actual deployment depth of the transmitter, the TL is 192.683 dB. As was noted above, the transmitter level in LOAPEX (Ref. 7) was about 194 dB. Thus, the axial wave could not have been detected for this configuration (range of 3 200 km and transmitter depth of 500 m).

The distances between the sound channel axis depth and the actual transmitter deployment depths were similar at stations T2300 and T1600 (328.1 and 421.4 m for transmitter depths of 500 and 350 m, respectively); hence, calculations similar to that shown in Fig. 7 indicate that the axial waves radiated by the transmitter deployed at stations T2300 and T1600 could also not have been detected.

C. Dependence on frequency

One transmitted M-sequence is well approximated by a signal with center frequency of 75 Hz and bandwidth of 37.5 Hz. (This is the full width between half-amplitude points in the transform envelope.) Figure 8 shows the dependence of axial wave TL (at a single frequency) when the transmitter is at the sound channel axis at station T1000 (i.e., at range 990 km.) The receiver depth is the channel axis depth at the VLA. The frequencies are taken to be 56.25, 75, and 93.75 Hz. Axial wave TL is seen to increase with decreasing frequency (i.e., the amplitude decreases with decreasing frequency.) This result depends on the number of overlapping cusped caustic neighborhoods. At lower frequency, these neighborhoods are spatially larger, and more such neighborhoods overlap, and hence more waves interfere with the wave propagating along the axis. For this example (at 990 km), 16 waves interfere at 56.25 Hz, 14 at 75 Hz, and 12 at 93.75 Hz.

To investigate variation in the time domain, it is natural to define the axial wave propagation time, T, as

$$T = \frac{\partial}{\partial \omega}\tilde{\varphi}, \tag{1}$$

where ω is a cyclic frequency and $\tilde{\varphi}$ is a phase function of the axial wave $\tilde{H}$ given by Eqs. (A14) and (A15),

$$\tilde{H} = \exp(i\tilde{\varphi})|\tilde{H}|. \tag{2}$$

Table I shows the arrival time of the axial wave and the arrival time of the lowest mode in the adiabatic approximation. The difference in arrival times decreases as the frequency increases. At 75 Hz, the difference is 36.8 ms. By comparison, the arrival time of a fictitious wave traveling along the axis—the reference arrival time—is 669.483 s and does not depend on frequency.

When at fixed frequency the propagation range decreases, the difference in arrival time between the axial wave and the lowest mode in the adiabatic approximation de-

TABLE I. Arrival times of the axial wave, t_{aw} and the lowest mode in the adiabatic approximation t_{lm} at frequency f with the transmitter at the sound channel axis. The receiver is at range 990 km and the depth of the sound channel axis (635.678 m) at the location of the VLA.

f (Hz)	t_{aw} (s)	t_{lm} (s)	$t_{aw}-t_{lm}$ (ms)
56.25	669.4511	669.4134	37.7
75.00	669.4620	669.4252	36.8
93.75	669.4657	669.4438	21.9

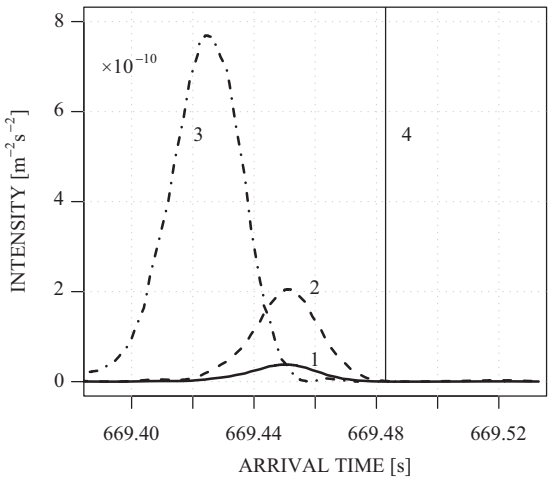


FIG. 9. Predicted intensity envelopes of the axial wave at 990 km for the smooth range-dependent ocean. The receiver is at the channel axis, and the transmitter is at station T1000. Curve (1) corresponds to a transmitter depth of 800 m and (2) the transmitter depth at the channel axis. Curve (3) is the intensity of the lowest mode in the adiabatic approximation when the depth of the transmitter is 800 m. Vertical line (4) indicates the reference arrival time for the wave traveling along the axis.

creases as well. Hence, all of the following calculations will be made for station T1000, where this difference is largest.

To calculate numerically the time domain representation of the axial wave, one considers the Fourier transform over the truncated domain $[\omega_{\rm lo}, \omega_{\rm hi}]$ of the function $\tilde{H}(\omega)$ [see Eqs. (A14) and (A15)] weighted by an appropriate spectral weighting function $\tilde{F}(\omega)$:

$$H(t) = \frac{1}{2\pi}\int_{\omega_{\rm lo}}^{\omega_{\rm hi}} \tilde{F}(\omega)\tilde{H}(\omega)e^{i\omega t}d\omega. \tag{3}$$

As noted above, an M-sequence is well approximated (after pulse compression) by a pulsed sinusoid with a center frequency of 75 Hz modulated by a Gaussian envelope with bandwidth of 37.5 Hz. In this case

$$\tilde{F}(\omega) = \exp\left(-\tfrac{1}{2}\sigma_t^2(\omega-\omega_0)^2\right), \tag{4}$$

where

$$\omega_0 = 150\pi\ \text{s}^{-1}, \quad \sigma_t = \frac{\sqrt{2\log 2}}{37.5\pi}, \quad s = 0.009\,994\ \text{s}. \tag{5}$$

The choice of integration interval $[\omega_{\rm lo}, \omega_{\rm hi}]=[\omega_0-\Delta_\omega, \omega_0+\Delta_\omega]$, with $\Delta_\omega=75\pi\ \text{s}^{-1}$, gives

$$\int_{\omega_0-\Delta_\omega}^{\omega_0+\Delta_\omega} \tilde{F}^2(\omega)d\omega \Big/ \int_{-\infty}^{+\infty} \tilde{F}^2(\omega)d\omega = 0.999\,132. \tag{6}$$

This integration interval is twice the bandwidth. Thus, the use of this interval $2\pi[37.5, 112.5]$ rad/s ([37.5, 112.5] Hz) provides adequate accuracy for pulse propagation modeling.

Figure 9 shows the time domain intensity envelopes of the axial wave for a receiver on the sound channel axis. The propagation range is 990 km. The maximal values of these envelopes are −120.273 and −112.941 dB for transmitter depths of 800 m and the channel axis, respectively. The maximal value of the intensity envelope of the lowest mode in the adiabatic approximation, $|H_{\rm lm}(t)|^2$, for a transmitter depth of 800 m is −107.200 dB. For comparison, the arrival time, or reference time, of the intensity peak of a wave traveling along the axis is shown: if this wave could be described by geometrical acoustics formulas (valid except at some ranges where its maximum value tends to infinity), the intensity envelope would have a shape similar to that of the axial wave. The reference time is 669.483 s. At the ranges under consideration, however, the wave traveling along the axis would interfere with other waves corresponding to rays with small launch angles and therefore cannot be described by simple geometrical acoustics formulas.

Figure 9 shows a 27 ms delay between the axial wave arrival and the lowest mode arrival. In this case (the delay is of the same order as the pulse width), the delay would be marginally adequate to resolve the axial wave were the two pulses of equal size, but the much larger modal signal will dominate and mask the axial wave signal for even the most favorable transmitter location (i.e., at the channel axis).

D. Effect of ocean internal waves on the axial wave

To study the effect of the environmental variability on the axial wave, a simple (but realistic) model for sound speed perturbations due to internal waves (see Appendix B) is used. Simulations were based on the buoyancy frequency profile based on LOAPEX CTD measurements made at station T50. To simplify the description of the sound speed perturbations, a single frequency corresponding to the semi-diurnal tide was chosen, and only the first ten modes were utilized. These modes introduce ocean structure with horizontal wavelengths ranging from 150 km down to 15 km. The complete environment is then formed by adding the smooth approximation—which now assumes the role of the "background"—and the (zero-mean) perturbations.

Simulations of the axial wave with this ocean model indicate that the maximal value of the intensity envelope is essentially zero for "full strength" perturbations. To understand the nature of this result, calculations were made for a set of ocean models in which the sound speed perturbations due to internal waves are scaled by an arbitrary small coefficient ε. This coefficient ε is taken to be 0.075, 0.09, and 0.15. (Results in the previous sections correspond to $\varepsilon=0.0$.)

Figure 10 shows the time domain intensity envelope of the axial wave, $|H(t)|^2$, at the range of 990 km, when the transmitter and receiver are placed at the channel axis of the unperturbed ocean for different values of ε. For $\varepsilon=0.075$, the maximal intensity is −124.200 dB; for $\varepsilon=0.09$, it is −126.559 dB; and for $\varepsilon=0.15$, it is −136.001 dB. For comparison, the maximal intensity in an ocean without internal waves is −112.941 dB (see Fig. 9). For a more realistic full strength value of ε of 1, the internal waves essentially randomize the depth of the sound channel axis to such an extent that they completely destroy the coherent structure of the axial wave, rendering it undetectable.

IV. SUMMARY

A ray diagram for a transmitter located at the sound channel axis in an ocean with a typical mid-latitude sound

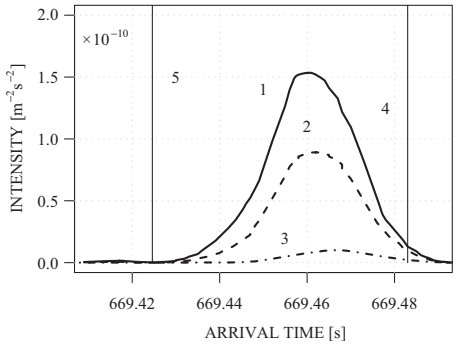


FIG. 10. Predicted axial wave time domain intensity at 990 km range. Transmitter and receiver are at the sound channel axis of the unperturbed ocean. The strength of the sound speed fluctuations is parametrized by a multiplicative coefficient ε: (1) corresponds to $\varepsilon=0.075$, (2) to $\varepsilon=0.09$, and (3) to $\varepsilon=0.15$. Vertical line (4) indicates the reference arrival time of the wave traveling directly along the sound channel axis, and vertical line (5) shows the arrival time of the lowest mode in the adiabatic approximation, both for an unperturbed ocean.

speed profile is characterized by the formation of numerous repeated cusped caustics along the axis. In neighborhoods of these cusped caustics, geometrical acoustics approximations are not valid because their application presupposes that the waves associated with individual ray paths do not interfere with one another. Neighborhoods of interference grow with range, and at a certain propagation range they overlap. This results in the formation of a diffractive (as opposed to ray, i.e., geometrical acoustics) component of the field—the axial wave—that propagates along the axis.

Axial wave propagation modeling has been carried out here for the environment realized during the 2004 LOAPEX experiment, which involved late-summer conditions in the North Pacific.

The two-dimensional range-dependent sound speed field was modeled to a first approximation by smooth polynomial and spline interpolation of the actual sound speed profiles derived from CTD measurements obtained along the 3200 km propagation path. A second, more complex, approximation was constructed by adding to the first approximation a zero-mean sound speed fluctuation field representing sound speed perturbations due to a (simplified) Garrett–Munk internal-wave field.

With respect to range, the TL of the axial wave was seen to be worse than cylindrical spreading. As an example, for 68.2 Hz (one of the LOAPEX frequencies), the smooth sound speed field and an optimal geometry (transmitter and receiver on the channel axis), the axial wave loses 117–120 dB over 3200 km. Given that the LOAPEX transmitter level was about 194 dB, the axial wave might have been detectable over ambient noise assuming some form of period averaging or other post-processing gain.

Actual LOAPEX transmitter deployment depths were not on the channel axis. With respect to transmitter depth, the axial wave strength was shown to decrease with increasing distance from the channel axis. At stations T3200, T2300, and T1600, these distances were at least 362.5, 328.1, and 421.4 m, respectively. In all these cases, the transmitter was too far off-axis to launch a detectable axial wave. Configurations were more favorable at stations T1000, T500, and T250, where separations were 125.9, 119.5, and 117.9 m, respectively.

With respect to frequency, the axial wave strength was seen to increase with frequency. This is related to the number of interfering cusped caustic regions, which decrease as frequency increases, i.e., as wavelengths get shorter. The corresponding propagation time was examined for station T1000, where the range was 990 km. At this station, the transmitter carrier frequency was 75 Hz. The axial wave was shown to arrive later than the lowest mode in the adiabatic approximation by more than a carrier period and hence might be separately resolvable.

To formally determine resolvability, it is necessary to incorporate bandwidth information. The axial wave time domain signature was synthesized by inverse Fourier transforming the axial wave function weighted by a Gaussian with bandwidth of 37.5 Hz centered at 75 Hz. (These values approximate the spectral shape of a LOAPEX 75 Hz M-sequence.) At 990 km, the axial wave pulse was shown (for a transmitter depth of 800 m or at the axis, and the smooth sound speed field) to be dominated by the nearby (earlier) pulse from the lowest mode in the adiabatic approximation to the extent that even with a delay of 27 ms, the axial wave pulse would not be separately resolvable. Resolution capacity degrades further for shorter ranges because the difference in arrival times between the two pulses decreases.

Predictions for the more complex—and hence more realistic—ocean model indicated that randomization of the sound speed seriously degrades the coherent structure of the axial wave. These predictions imply that it would not be possible to detect the axial wave for any transmitter/receiver configuration for the environmental settings corresponding to those encountered during LOAPEX.

As is well known, the effect of internal waves is negligible in the Arctic Ocean. This region is characterized by the formation of a near-surface waveguide where the axial wave exists. This situation coupled with the weakness of the internal-wave field might possibly create favorable conditions for axial wave detection.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research through the Ocean Acoustics Program Grant No. N00014-06-1-0421 and by the Office of Naval Research Global through Grant No. N00014-08-1-4013.

APPENDIX A: INTEGRAL REPRESENTATION OF THE AXIAL WAVE

In the frequency domain in a range-independent ocean where the variables are separable, the formula for the axial wave is obtained by transforming the contour integral that gives the exact solution of the point transmitter problem. The integral representation of the axial wave is of a different form for a normal waveguide[9] where the wave traveling along the sound channel axis is the last to arrive and for the abnormal waveguide where the situation is opposite. In most

long-range propagation experiments (in particular, in LOAPEX), deep-water waveguides are normal. Therefore, only normal waveguides will be considered below.

The formula for the axial wave may be written in terms that can be obtained for a range-dependent ocean as well.[5] This formula includes the solutions to the Helmholtz equation

$$\frac{\partial^2 u^0}{\partial x^2} + \frac{\partial^2 u^0}{\partial z^2} + \frac{\omega^2}{\bar{c}^2(z)} u^0 = 0, \tag{A1}$$

where ω is a cyclic frequency and $c(z)$ is the sound speed that can be expanded in the case of small $|z-z_0|$ in the form

$$\bar{c}(z) = \bar{c}_0 + \frac{\bar{c}_2}{2!}(z-z_0)^2 + \frac{\bar{c}_3}{3!}(z-z_0)^3 + \cdots, \quad \bar{c}_2 > 0. \tag{A2}$$

These solutions are concentrated near the sound channel axis $z=z_0$ and decay exponentially outside a narrow strip containing the axis as ω tends to infinity. They are derived in the form of a product of the exponential function and the parabolic cylinder function[10,11] $D_q(\Psi)$, $q=0,1,\ldots$,

$$u^0(x,z) = u_q^0(x,z) = \exp\left\{ i\omega \sum_{m=0}^{M+1} \alpha_m^0(x,\xi)\omega^{-m/2} + O(\omega^{-M/2}) \right\} D_q\left(\sum_{m=0}^{M-1} \beta_m^0(x,\xi)\omega^{-m/2} + O(\omega^{-M/2}) \right). \tag{A3}$$

The coefficients $\alpha_m^0(x,\xi)$ and $\beta_m^0(x,\xi)$, which are polynomials in $\xi=\sqrt{\omega}\,(z-z_0)$, are found from a recursive system of partial differential equations resulting from the substitution of Eq. (A3) into Eq. (A1). M is taken equal to 3, which gives a residual [the difference between the left-hand side of Eq. (A1) and zero] of order $O(\omega^{1/2})$. As a result, one gets[5]

$$u_q^0(x,z) = \exp\left\{ i\left(\frac{\omega}{\bar{c}_0}x - \left(q+\frac{1}{2}\right)\frac{x}{a} + \frac{i}{4\sqrt{3}}\chi\frac{\xi}{a}\frac{1}{\sqrt{\omega}} + \left[\frac{i}{8}\left(b-1+\frac{5}{6}\chi^2\right)\frac{\xi^2}{a^2} - \frac{1}{2}\frac{\bar{c}_0}{a^2}\left(b\left(q+\frac{1}{2}\right)^2 + \frac{1}{4}\left(b-1+\frac{3}{4}\chi^2\right)\right)x \right]\frac{1}{\omega} + O\left(\frac{1}{\omega^{3/2}}\right)\right)\right\} D_q\left(\sqrt{2}\left(\frac{\xi}{\sqrt{a\bar{c}_0}} + \frac{1}{\sqrt{3}}\frac{\sqrt{\bar{c}_0}}{\sqrt{a}}\chi\left[\frac{1}{4a\bar{c}_0}\xi^2 + \left(q+\frac{1}{2}\right)\right]\frac{1}{\sqrt{\omega}} + \frac{1}{12}\left[\frac{1}{a^{5/2}\sqrt{\bar{c}_0}}(b-1+\chi^2)\xi^3 + 3\frac{\sqrt{\bar{c}_0}}{a^{3/2}}\left(q+\frac{1}{2}\right)\left(b-1+\frac{1}{3}\chi^2\right)\xi\right]\frac{1}{\omega} + O\left(\frac{1}{\omega^{3/2}}\right)\right)\right), \tag{A4}$$

where $\bar{c}_0$ is the axial sound speed, $a=\sqrt{\bar{c}_0/\bar{c}_2}$ is the characteristic dimension of the waveguide, the value

$$\chi = \frac{2}{3^{3/2}}\frac{\bar{c}_3\sqrt{\bar{c}_0}}{\bar{c}_0^{3/2}} \tag{A5}$$

characterizes the difference of the actual waveguide from the appropriate symmetric waveguide, where $\chi=0$, and

$$b = \frac{1}{8}\left(\frac{\bar{c}_4\bar{c}_0}{\bar{c}_2^2} - 1 - \frac{5}{3}\frac{\bar{c}_3^2\bar{c}_0}{\bar{c}_2^3}\right) \tag{A6}$$

defines the type of a waveguide. If $b<0$ the waveguide is normal, and if $b>0$ it is abnormal.

Saying that ω tends to infinity means that the dimensionless frequency $\kappa=(\omega/\bar{c}_0)\,a$ is large. In LOAPEX, for the sound speed profile shown in Fig. 4 at the frequency of 75 Hz one obtains $\kappa=845.87$.

In a normal range-independent waveguide, the integral representation of the axial wave, $\tilde{H}^0(\omega)=\tilde{H}^0(x_r,z_r,z_s;\omega)$ has the following form:

$$\tilde{H}^0(\omega) \sim \frac{1}{\sqrt{ax_r}} \int_{\infty\exp(-3i\pi/4)}^{\infty\exp(i\pi/4)} p^0(\tau) u^0_{\tau-1/2}(0,z_s) \times u^0_{\tau-1/2}(x_r,z_r)\,d\tau. \tag{A7}$$

Here $(0,0,z_s)$ and $(x_r,0,z_r)$ are the coordinates of the transmitter and receiver, respectively, $z_s \leqslant z_r$; $u^0_{\tau-1/2}(x,z)$ is obtained from $u^0_{\tau-1/2}(x,z)$ by replacing $D_{\tau-1/2}(\Psi)$ with $D_{\tau-1/2}(-\Psi)$. The weight function

$$p^0(\tau) = \frac{e^{-i\pi/4}}{4\pi}\exp\left(-i\frac{\pi}{2}(K^0-1)\right)2^{-\tau} \times \frac{\Gamma(\tau+1/2)}{\Gamma^2(\tau/2+3/4)}\frac{e^{i\pi K^0\tau}}{1+ie^{i\pi\tau}}[D_{\tau-1/2}(0)]^{-2} \tag{A8}$$

includes the integer K^0 that satisfies the inequalities

$$\frac{x_r}{\pi a}\left(1-|b|^{1/2}\left(\frac{\omega}{\bar{c}_0}x_r\right)^{-0.5+\delta}\right) \leqslant K^0 < \frac{x_r}{\pi a}\left(1-|b|^{1/2}\left(\frac{\omega}{\bar{c}_0}x_r\right)^{-0.5+\delta}\right)+1, \tag{A9}$$

where $0<\delta<1/2$. With such a choice of the number K^0, the wave corresponding to a ray that crosses the waveguide axis K^0-1 times will be the last wave that does not interfere with

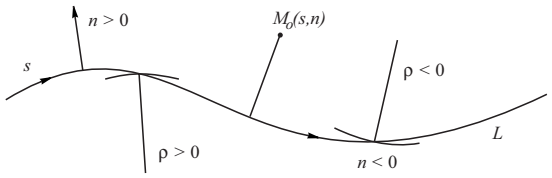


FIG. 11. The (s,n)-coordinates near the waveguide axis L in the (x,z)-plane.

the wave propagating directly along the waveguide axis and can be observed independently of the latter.[5] The selection of a constant δ is discussed in Ref. 5. For most medium models, δ can be taken to be 0.1. In all our simulations of the axial wave, we put δ equal to 0.1 as well.

If $z_r < z_s$, the integral representation of the axial wave takes on the form

$$\tilde{H}^0(\omega) \sim \frac{1}{\sqrt{ax_r}} \int_{\infty \exp(-3i\pi/4)}^{\infty \exp(i\pi/4)} p^0(\tau) u^0_{\tau-1/2}(0,z_s) \times \underline{u}^0_{\tau-1/2}(x_r,z_r) d\tau. \quad \text{(A10)}$$

The integration contour in Eqs. (A7) and (A10) begins at infinity in the complex τ-plane and goes to $\tau=0$ along the straight line arg $\tau=-3\pi/4$. From $\tau=0$ it goes out to infinity along the straight line arg $\tau=\pi/4$.

In a range-dependent ocean it is assumed that the ocean sound speed, c, depends on depth coordinate z and coordinate x measured along the transmitter-receiver direction. In this case the sound channel axis is some curve L. Assume that the curve L can be represented parametrically as $x=x(s)$, $z=z(s)$, where the parameter s is taken to be the arc length along L measured in a given direction from some initial point. Introduce a new system of coordinates (s,n) in the vicinity of L. The coordinates (s,n) of an arbitrary point M_0 will be defined as follows: $|n|$ is the length of the normal dropped from M_0 to L, and s is the arc length along L from the initial point to the base of the normal (Fig. 11). In addition, assume that n is positive if M_0 is to the left of L as one moves along L in the direction of increasing values of s. The Lamé coefficients h_1 and h_2 of the coordinate system (s,n) are

$$h_1 = 1 + \frac{n}{\rho(s)}, \quad h_2 = 1. \quad \text{(A11)}$$

Here $\rho(s)$ is the radius of curvature of L. It is assumed to be positive if the center of curvature lies to the right of L, i.e., among negative values of n, and the radius of curvature is negative otherwise (see Fig. 11).

The sound speed $c(x,z)$ in (s,n)-coordinates is denoted by $c(s,n)$ as before. It is assumed that the sound speed $c(s,n)$ is differentiable a sufficient number of times with respect to both variables s and n and therefore can be expanded in the case of small $|n|$ in the form

$$c(s,n) = c_0(s) + c_1(s)n + \frac{1}{2!}c_2(s)n^2 + \cdots. \quad \text{(A12)}$$

In a range-dependent ocean, the integral representation of the axial wave is constructed from the solutions to the Helmholtz equation written in (s,n)-coordinates. These solutions $u_q(s,n)$, $q=0,1,\ldots$, have the nature of waves propagating along the range-variable sound channel axis L and tending rapidly to zero outside a boundary layer surrounding the axis. They appear in a form similar to Eq. (A3),

$$u_q(s,n) = \exp\left\{ i\omega \sum_{m=0}^{M+1} \alpha_m(s,\nu)\omega^{-m/2} + O(\omega^{-M/2}) \right\} \times D_q\left(\sum_{m=0}^{M-1} \beta_m(s,\nu)\omega^{-m/2} + O(\omega^{-M/2}) \right). \quad \text{(A13)}$$

The coefficients $\alpha_m(s,\nu)$ and $\beta_m(s,\nu)$, which are polynomials in $\nu=\sqrt{\omega}n$, are obtained from a recursive system of partial differential equations resulting from the substitution of Eq. (A13) into the Helmholtz equation.[6] Introduction of the variable $\nu=\sqrt{\omega}n$ and representation of the desired solution in the form of a product of the rapidly varying (for large ω) exponential function and the parabolic cylinder function, which varies more slowly, means that in finding of the solution (A13), one uses the parabolic equation method (or the boundary layer method) pioneered by Tappert.[12]

As opposed to the adiabatic approximation of low-order modes that depend only on medium properties along the vertical line parallel to the z-axis, the newly obtained solutions to the Helmholtz equation accumulate the information about medium properties near the range-variable sound channel axis along the whole propagation range. The coefficients $\alpha_m(s,\nu)$ and $\beta_m(s,\nu)$ are found[6] up to M equal to 3. In this case, substituting the newly obtained solutions into the Helmholtz equation yields residuals of order $O(\omega^{-1/2})$.

Let us assume that in the coordinate system (s,n) introduced in the vicinity of the sound channel axis L on the (x,z)-plane, the transmitter is at $s=0$, $n=n_s$, and the receiver is at $s=s_r$, $n=n_r$. By analogy with Eq. (A7) for the axial wave in a range-independent medium, if $n_s \leqslant n_r$ the integral representation of the axial wave, $\tilde{H}(\omega)=\tilde{H}(s_r,n_r,n_s;\omega)$, in a range-dependent ocean will have the form

$$\tilde{H}(\omega) \sim \frac{1}{\sqrt{as_r}} \int_{\infty \exp(-3i\pi/4)}^{\infty \exp(i\pi/4)} p(\tau) \underline{u}_{\tau-1/2}(0,n_s) u_{\tau-1/2}(s_r,n_r) d\tau. \quad \text{(A14)}$$

where $p(\tau)$ is the weight function, $\underline{u}_{\tau-1/2}(s,n)$ is obtained from $u_{\tau-1/2}(s,n)$ by replacing $D_{\tau-1/2}(\Psi)$ with $D_{\tau-1/2}(-\Psi)$, and $a=\sqrt{c_0/c_2}$ is the characteristic dimension of the range-independent waveguide with the sound speed $\bar{c}(z)$ along the vertical line passing through the transmitter $c(0,z)=\bar{c}(z)$ [see also Eq. (A2)].

If $n_r < n_s$ the integral representation of the axial wave has the form

$$\tilde{H}(\omega) \sim \frac{1}{\sqrt{as_r}} \int_{\infty \exp(-3i\pi/4)}^{\infty \exp(i\pi/4)} p(\tau) u^0_{\tau-1/2}(0,n_s) \underline{u}^0_{\tau-1/2}(s_r,n_r) d\tau. \quad \text{(A15)}$$

In the case of a range-independent medium, the integrals (A14) and (A15) must reduce to the integrals (A7) and (A10), respectively. On the other hand, the weight function

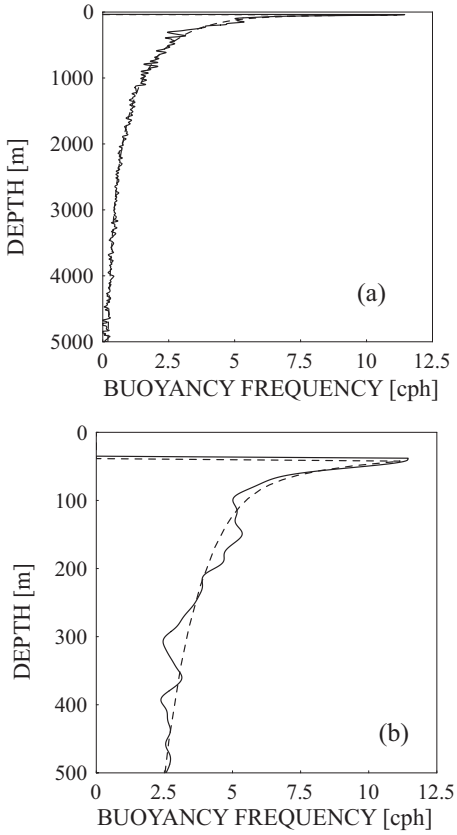


FIG. 12. Buoyancy frequency profile obtained from a CTD cast at station T50 (solid) and its smooth approximation (dashed). Panel (a) corresponds to the depth interval (0,5 000) m and panel (b) to the depth interval (0,500) m.

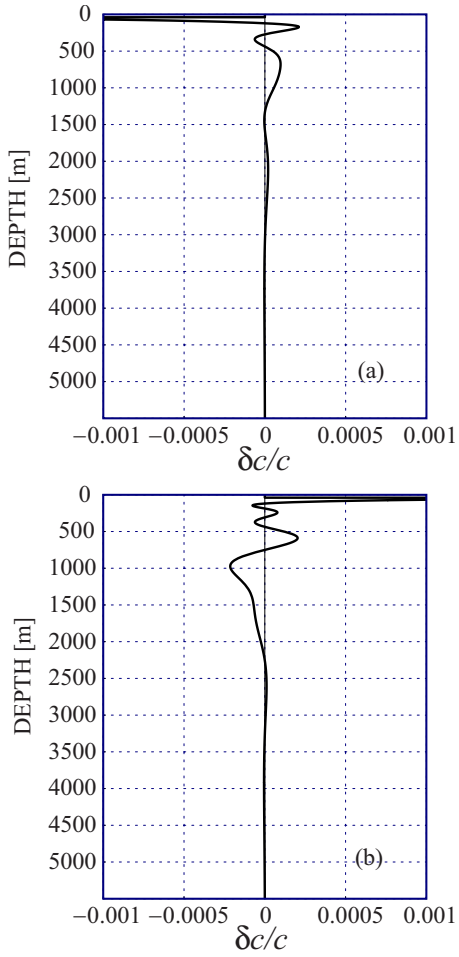


FIG. 13. (Color online) Sound-speed perturbations versus depth for one realization of a ten-mode internal-wave field. Panel (a) shows the perturbation at the VLA, and panel (b) shows the perturbation at 990 km from the VLA (station T1000).

$p(\tau)$ must be independent of medium properties. It can be shown that this requirement is equivalent to the localization principle,[13] which asserts that the high-frequency asymptotic behavior of a wave field depends only on properties of the medium in the vicinity of the ray. Thus, the weight function $p(\tau)$ in Eqs. (A14) and (A15) is given by the same expression (A8) as in the case of a range-independent ocean. The integer K contained in $p(\tau)$ satisfies the inequalities that are the generalization of Eq. (A9) to the case of a range-dependent medium.[6]

APPENDIX B: SOUND-SPEED PERTURBATIONS DUE TO INTERNAL WAVES

We shall assume that the ocean sound speed along the transmitter-receiver direction $\hat{c}(x,z,t)$, consists of the unperturbed sound speed profile $c(x,z)$ and a small perturbation, $\delta c(x,z,t)$, dependent on range x, depth coordinate z, and time l,

$$\hat{c}(x,z,t) = c(x,z) + \delta c(x,z,t). \tag{B1}$$

The vertical position z varies from $z=0$ at the ocean surface to $z=-D$ $(D>0)$ at the ocean bottom. The second horizontal spatial coordinate has been omitted since for the choice of δc below, the cross-range gradients of the sound speed are typically two orders of magnitude smaller than the vertical gradients, and therefore the sound propagation is believed to be sufficiently well described as two dimensional (i.e., having no out-of-plane scattering).

The quantity δc will represent the sound speed perturbations due to a field of internal waves. For these numerical experiments, δc is constructed from a superposition of linear internal waves, constrained by the following assumptions: the ocean depth $D=6$ km is range independent and the buoyancy profile $N(z)$ is obtained as a smooth approximation of measured data. Figure 12 shows the buoyancy frequency profile from the LOAPEX CTD cast at station T50 and a smooth approximation used for simulations.

Neglecting the small effects due to internal-wave currents, the sound speed perturbations due to internal waves are given by[14,15]

$$\delta c/c \approx GN^2(z)\zeta(x,z,t), \tag{B2}$$

where ζ is the vertical water displacement due to the internal waves, and G, assumed constant here,[15] has the value of 3.0 s^2/m.

The displacement ζ is constructed from the superposition of a subset field of linear internal waves traveling along the transmitter-receiver direction. For linear waves in a flat bottom ocean, the displacement ζ is given by

$$\zeta(x,z,t) = \mathrm{Re}(W(z)e^{i(kx-\Omega t)}), \tag{B3}$$

where $W(z)$ satisfies

$$(\Omega^2 - f_i^2)\frac{d^2W}{dz^2} + (N^2(z) - \Omega^2)k^2W = 0 \tag{B4}$$

subject to the boundary conditions $W(0)=W(-D)=0$. Here the inertial frequency f_i is taken to be 1 cycle/day $=(\pi/4.32)\times 10^{-4}$ s^{-1}, corresponding to a latitude of 30°. Given the wave frequency Ω there can be found a sequence of modes $W_j(z)$ and corresponding wave numbers k_j, $j=1,2,\ldots$, that solve Eq. (B4). These modes form an orthogonal set so that

$$\int_{-D}^{0} (N^2(z) - f_i^2)W_m(z)W_n(z)dz = \gamma\delta_{mn}, \tag{B5}$$

where γ is any arbitrary normalization constant independent of m and n; for later convenience[15] γ is taken to be $2/(BN_0^2)$. Here $B=1$ km and N_0 is found from the condition[16]

$$N_0B = \int_{-D}^{0} N(z)dz. \tag{B6}$$

To simplify the description of the sound speed perturbations, a single frequency Ω is chosen, $\Omega=2\pi/12.5$ rad h^{-1} $=(\pi/2.25)\times 10^{-4}$ s^{-1}, corresponding to the semi-diurnal tide. Summing over the first ten modes associated with that tidal frequency produces the model internal-wave field or, more specifically, an internal-tide field.

After summing, the internal-wave displacement is given by

$$\zeta(x,z,t) = \mathrm{Re}\left(\sum_{j=1}^{10} A_jW_j(z)e^{i(k_jx-\Omega t+\lambda_j)}\right), \tag{B7}$$

where λ_j are random phases. $A_je^{i\lambda_j}$ is a complex random variable with A_j weighted according to

$$\langle A_j^2\rangle = q(j^2 + j_*^2)^{-1}, \tag{B8}$$

where $j_*=3$. If the Garrett–Munk[17] spectral density, a function of mode number and wave number, is integrated over wave number k, then the remaining dependence of mode number j is as above.[15] Following Ref. 14, let

$$q = \frac{E_0 j_*}{\pi}\frac{BN_0^2}{2\hat{\rho}}, \tag{B9}$$

where $E_0=0.4$ J/cm$^2=4\times 10^3$ kg/s^2 is the mean energy and $\hat{\rho}=1$ g/cm$^3=10^3$ kg/m^3 is the density of water.

Construction of the internal-wave modes $W_j(z)$ follows the method proposed by Colosi and Brown,[16] where Wentzel–Kramers–Brillouin scaling of both the modal amplitude and the depth coordinate is used,

$$V(z) = (N(z)/N_0)^{1/2}W(z) \tag{B10}$$

and

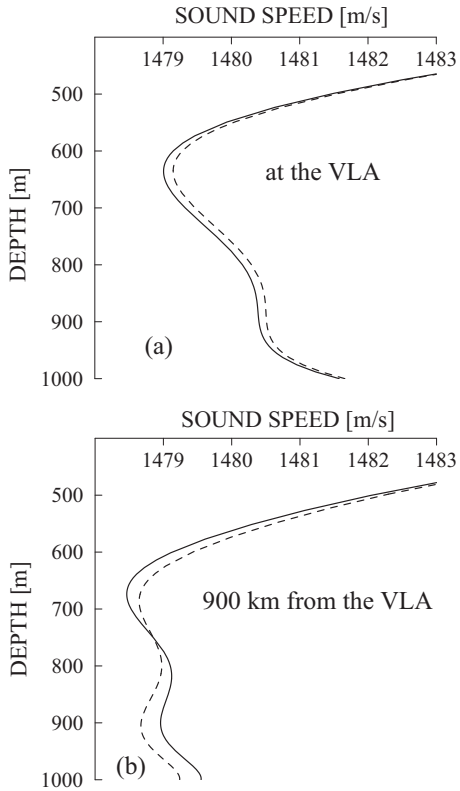


FIG. 14. Sound-speed profiles without internal waves (solid) and for one realization of a ten-mode internal-wave field (dashed). Panel (a) shows these profiles at the VLA and panel (b) shows these at 990 km from the VLA.

$$\hat{\xi}(z) = \frac{1}{N_0B}\int_{-D}^{z} N(z')dz'. \tag{B11}$$

As a result,[16] the approximate solution of Eq. (B4) satisfying the boundary conditions $W(0)=W(-D)=0$ and the normalizing condition (B5) will be of the form

$$W_j(z) = \frac{\sqrt{2}\sqrt{N(z)}}{N_0^{3/2}\sqrt{B}}\sin\left(\frac{k_jBN_0}{\sqrt{\Omega^2 - f_i^2}}\hat{\xi}(z)\right), \tag{B12}$$

with

$$\frac{k_jBN_0}{\sqrt{\Omega^2 - f_i^2}} = j\pi, \quad j = 1,2,\ldots. \tag{B13}$$

For the buoyancy frequency profile shown in Fig. 12, $B_0=0.008\,97$ s$^{-1}\approx 5.16$ cycles/h. Figure 13 shows the sound speed fluctuations induced by the first ten modes of the internal-wave field. Figure 14 demonstrates the perturbed and unperturbed sound speed profiles at two propagation ranges: at the VLA and at station T1000. Modes 1–10 introduce the ocean structure with horizontal wavelengths ranging from 150 down to 15 km.

[1]T. Pearcey, "The structure of an electromagnetic field in the neighborhood of a cusp of a caustic," J. Acoust. Soc. Am. **37**, 311–317 (1946).
[2]Yu. A. Kravtsov and Yu. I. Orlov, *Geometrical Optics of Inhomogeneous Media* (Springer-Verlag, New York, 1991).
[3]V. S. Buldyrev, "The field of a point transmitter in a waveguide," Tr. Mat. Inst. Steklova **115**, 78–102 (1971).
[4]V. S. Buldyrev, A. I. Lanin, and Z. A. Yanson, "Calculation of a field at

the axis of a symmetric waveguide," Voprosy Dynam. Teor. **14**, 84–93 (1974).
[5]N. S. Grigorieva and G. M. Fridman, "Axial wave in long-range propagation in a range-independent ocean," J. Comput. Acoust. **12**, 127–147 (2004).
[6]N. S. Grigorieva and G. M. Fridman, "Effect of horizontal inhomogeneity of the ocean on interference of near-axial waves in long-range acoustic propagation," J. Comput. Acoust. **14**, 415–443 (2006).
[7]J. A. Mercer, R. K. Andrew, B. M. Howe, and J. A. Colosi, "Cruise report: Long-range Ocean Acoustic Propagation Experiment (LOAPEX)," Appl. Phys. Lab., University of Washington, 2005.
[8]N. S. Grigorieva and G. M. Fridman, "Dependence of the axial wave on range and sound speed properties in a range-independent ocean," J. Comput. Acoust. **13**, 259–278 (2005).
[9]W. H. Munk, P. F. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, London, 1995).
[10]H. Bateman and A. Erdélyi, *Higher Transcendental Functions*, (McGraw-Hill, New York, 1953), Vol. **2**.
[11]*Handbook of Mathematical Functions*, edited by M. Abramovitz and I. Stegun (Dover, New York, 1970), Chap. 19, pp. 685–700.
[12]F. D. Tappert, "The parabolic approximation method," in *Wave Propagation and Underwater Acoustics*, edited by J. B. Keller and J. S. Papadakis (Springer-Verlag, Berlin, 1977), pp. 224–284.
[13]V. M. Babič and V. S. Buldyrev, *Short-Wavelength Diffraction Theory* (Springer-Verlag, New York, 1991), p. 38.
[14]S. M. Flatté, R. Dashen, W. H. Munk, K. M. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, London, 1979).
[15]J. Simmen, S. M. Flatté, and G.-Yu. Wang, "Wavefront folding, chaos, and diffraction for sound propagation through ocean internal waves," J. Acoust. Soc. Am. **102**, 239–255 (1997).
[16]J. A. Colosi and M. G. Brown, "Efficient numerical simulation of stochastic internal-wave-induced sound speed perturbation field," J. Acoust. Soc. Am. **103**, 2232–2235 (1998).
[17]C. Garrett and W. H. Munk, "Space-time scales of internal waves: A progress report," J. Geophys. Res. **80**, 291–297 (1975).

Effects of incident field refraction on scattered field from vertically extended cylindrical targets in range-dependent ocean waveguides

Elizabeth T. Küsel and Purnima Ratilal
Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115

(Received 9 September 2008; revised 26 January 2009; accepted 27 January 2009)

The effect of incident field refraction on the scattered field from vertically extended cylindrical targets is investigated. A theoretical model for the total scattered field from a cylindrical target in a range-dependent ocean waveguide is developed from Green's theorem. The locally scattered field on the target surface is estimated as a function of the incident field by applying the appropriate boundary conditions on continuity of acoustic pressure and normal velocity, making the model applicable to general penetrable cylinders. The model can account for depth dependence in medium sound speed and hence refraction in the incident field along the target depth. Numerical implementation is done for a passive acoustic reflector, a long cylindrical air-filled rubber hose, often deployed vertically in experiments to provide calibration and charting consistency for wide-area active sonar systems. Analysis with the model indicates that refraction in the incident field along the target depth must be taken into account to accurately estimate the scattered field level from vertically extended cylindrical targets. It is demonstrated that the standard Ingenito waveguide target scattering model, which assumes that the incident field is planar along the target extent, can lead to significant errors of 10 dB or more in estimating the scattered field level.
 [DOI: 10.1121/1.3083235]

PACS number(s): 43.30.Gv, 43.20.Fn [WLS] Pages: 1930–1936

I. INTRODUCTION

In real ocean waveguides, the water column sound speed profile can vary significantly as a function of depth leading to acoustic field propagation that is highly refractive. Here, we investigate the effect of non-planar incident field on the scattered field from finite vertically extended cylindrical targets in range-dependent ocean waveguides. To better quantify the effect, we develop a theoretical model for the scattered field from a vertically extended cylindrical target that accounts for non-planar incident field over the target depth by direct application of Green's theorem.[1,2] The locally scattered field on the target surface at each depth is estimated as a function of the incident field by applying the boundary conditions on continuity of acoustic pressure and normal velocity, making the model applicable to general penetrable cylinders. Since the scattering contributions are calculated at each depth of the target, the object does not need to be located in an iso-speed layer. Furthermore, the scattering model for vertically extended cylindrical targets developed here can be implemented using both normal modes and the parabolic equation (PE) acoustic propagation models making the approach applicable to general range-dependent environments.[3,4]

The formulation is applied to calculate the scattered field from a cylindrical target[5] manufactured by BBN Systems and Technologies (Cambridge, MA) in several shallow water environments with different sound speed profile and bathymetry. The BBN target is often deployed during acoustic experiments[6,7] serving as a passive acoustic reflector for its known high target strength. Experiments that have deployed the BBN target include the ONR-sponsored 2006 Ocean Acoustic Waveguide Remote Sensing (OAWRS) Experiment at Georges Bank in the Gulf of Maine, and the 2001[8,9] and 2003 (Ref. 10) Acoustic Clutter Imaging Experiments in the New Jersey Strataform under the ONR Geoclutter Program. The BBN target is essentially a 30 m long and 7 cm diameter air-filled cylindrical hose made of gum rubber, suspended vertically off the seafloor at specified depths using a combination of anchors and floats. It provides a means to validate full-field waveguide scattering models and to minimize charting errors in range and bearing for active sonar systems. The scattered field from the BBN targets can also be compared to that from other targets of interest, such as fish schools, underwater vehicles, and the sea bottom, in order to determine their relative target or scattering strengths.

The scattering of sound by an object in an ocean waveguide is complicated due to multi-modal propagation and dispersion. The ocean-acoustic standard, Ingenito[11] waveguide target scattering model, is a widely used method[12–15] to calculate the single scattering from arbitrary sized objects in a horizontally stratified ocean waveguide. This approach is based on normal mode theory and it decomposes the incident field at the target center into modal plane waves. The object's plane wave scatter function is used to couple the incident and scattered modes which are then propagated to the receiver. The Ingenito model is therefore valid for large scatterers, large compared to the acoustic wavelength, in an ocean waveguide. However, it only applies to objects contained within an iso-speed layer because of modal plane wave decomposition at the target center. The Ingenito model cannot account for changes in medium sound speed along the target depth and assumes that the incident field is planar over the

 0001-4966/2009/125(4)/1930/7/$25.00

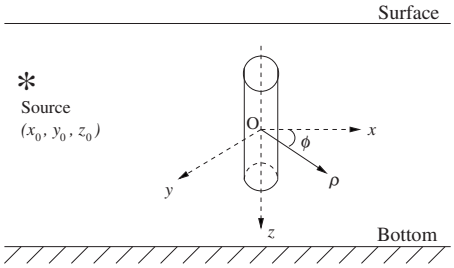


FIG. 1. Geometry of the scattering problem from a vertically extended cylindrical target of radius a and length L. The origin of the coordinate system is at the center of the cylinder.

target extent. Other approaches[16–22] for modeling object scattering in a waveguide are similarly limited, applicable to or implemented only for objects contained in an iso-speed layer. Some models[16,17,23–26] are only valid for horizontally stratified range-independent environments while others, such as the sonar equation,[27] are only valid for compact scatterers,[28] small compared to the acoustic wavelength, in a waveguide.

The theoretical vertically extended cylindrical target waveguide scattering (VETWS) model developed here is first calibrated against the Ingenito[11] model for the BBN target in a Pekeris waveguide where the target is contained within the iso-speed water column layer. The two models lead to scattered field levels as a function of range that match perfectly in this environment. The VETWS model is then implemented in a waveguide with the highly refractive water-column sound speed profile from the Gulf of Maine. In this environment, the BBN target extends over a depth where the water-column sound speed cannot be approximated as a constant and the field incident on the target is highly nonplanar. We show that when the Ingenito model is applied in this environment, it leads to significant errors of 10 dB or more in estimating the scattered field level for the BBN target.

Section II presents the theoretical approach and analytical formulation used to model the scattered field from a finite vertically extended cylindrical target in a general ocean waveguide. In Sec. III we describe the numerical implementation for the BBN target, and present results of calculations in both range-independent and range-dependent shallow water environments with different sound speed profiles. We compare the scattered field level using our model with that from the Ingenito model for each case.

II. THEORETICAL FORMULATION

In this section we develop the analytic formulation for the scattered field from a vertically extended cylindrical target in a range-dependent ocean waveguide. The origin of the coordinate system is placed at the target center with the z-axis pointing vertically downward and aligned with the axis of the cylinder, as shown in Fig. 1. The coordinates for points on the target are denoted by $\vec{r}_t=(x_t,y_t,z_t)=(\rho_t\cos\phi_t,\rho_t\sin\phi_t,z_t)$. Given a time harmonic source at location $\vec{r}_0$ and receiver at location $\vec{r}$, the total scattered field from the target is obtained by application of Green's theorem

$$P_{\text{scat}}(\vec{r})=-\oint_S [P(\vec{r}_t)\nabla_t G(\vec{r}|\vec{r}_t)-\nabla_t P(\vec{r}_t)G(\vec{r}|\vec{r}_t)]\cdot\vec{n}_t dA_t. \tag{1}$$

In Eq. (1), integration is performed over the surface S of the scatterer, $P(\vec{r}_t)$ is the total acoustic pressure on the target surface, $G(\vec{r}|\vec{r}_t)$ is the waveguide Green's function, and $\vec{n}_t$ is the unit vector normal to the surface of the object pointing away from the medium. Equation (1) is applicable to general targets with arbitrary boundary conditions.

The total pressure field on the surface of the target can be written as the sum of incident and scattered components

$$P(\vec{r}_t)=P_{\text{inc}}(\vec{r}_t)+P_{\text{scat}}(\vec{r}_t). \tag{2}$$

For a cylindrical target of radius $\rho_t=a$ and length L in an ocean waveguide, the local incident field on the surface of the target can be calculated numerically from an acoustic propagation model. The local scattered field on the surface of the cylindrical scatterer, on the other hand, is approximated in terms of Hankel functions as follows:

$$P_{\text{scat}}(\vec{r}_t)\approx\sum_{n=0}^{\infty}A_n(z_t)H_n^{(1)}(k(\vec{r}_t)\rho_t)\cos(n\phi_t). \tag{3}$$

At each depth point along the length of the target, this decomposition of the locally scattered field into cylindrical harmonics is valid since the acoustic mode cycle distances are typically much larger than the radius of the cylinder. In the above approximation, $H_n^{(1)}$ is the Hankel function of first kind and order n, and $k(\vec{r}_t)=\omega/c(\vec{r}_t)$ is the medium wavenumber, where ω is the circular frequency and $c(\vec{r}_t)$ is the range and depth-dependent medium sound speed. The depth-dependent coefficients $A_n(z_t)$ are optimally obtained by a least squares inversion applying the appropriate boundary condition at the surface of the target.

We assume that the radius of the cylinder is small compared to its length so that scattering is dominated by contributions from the curved surface of the cylinder and the contribution from the end caps is negligible. This also ensures that single scattering dominates for the vertically extended cylinder and multiple scattering with the waveguide boundaries is negligible.[29]

Next we examine three different boundary conditions for the finite vertically extended cylinder, the pressure-release, rigid, and penetrable surfaces.

A. Pressure-release target

The pressure-release condition states that the total pressure field at the target surface given by Eq. (2) must vanish. Substituting $P(\vec{r}_t)=0$ into Eq. (1) we observe that the first term under the integral sign goes to zero, and we obtain a simplified form for the total scattered field

$$P_{\text{scat}}(\vec{r})=\oint_S [\nabla_t P(\vec{r}_t)G(\vec{r}|\vec{r}_t)]\cdot\vec{n}_t dA_t. \tag{4}$$

In order to solve Eq. (4), we need to define the gradient of the local incident and local scattered pressure fields at the target surface. Over the curved surface of the circular cylinder,

$$\nabla_t P(\vec{r_t}) \cdot \vec{n_t} = -\frac{\partial P(\vec{r_t})}{\partial \rho_t}, \tag{5}$$

i.e., the gradient of the field dotted with the normal vector to the surface is just the derivative in the radial direction. The gradient of the incident field on the surface of the target can be calculated numerically from an acoustic propagation model. In order to calculate the gradient of the locally scattered field on the surface of the target, however, we first make use of Eq. (3) and the boundary condition to find the coefficients $A_n(z_t)$ and hence the local scattered field.

By substituting Eq. (3) and the pressure-release boundary condition into Eq. (2), we can write the resulting equation as a system of linear equations of the form

$$B\vec{q} = \vec{p} \tag{6}$$

at each depth point of the cylinder, where B is an $M \times N$ matrix of coefficients $H_n^{(1)}(k(\vec{r_t})\rho_t)\cos(n\phi_t)$, where M is the number of discrete points in azimuth around the cylinder and N is the number of terms of the Hankel function, $\vec{q}$ is the vector of coefficients $A_n(z_t)$ of length N we are interested in finding, and $\vec{p}$ is of size M and corresponds to $-P_{\text{inc}}(\vec{r_t})$, the negative incident pressure field at discrete azimuthal points around the cylinder. At each depth z_t of the target, we calculate the least-squares solution for $A_n(z_t)$ that minimizes $(\vec{p} - B\vec{q})$.

Once the coefficients $A_n(z_t)$ are determined and hence an approximation for the local scattered field, we use Eqs. (3) and (5) to calculate the gradient of the field as follows:

$$\frac{\partial P_{\text{scat}}(\vec{r_t})}{\partial \rho_t} = \sum_{n=0}^{\infty} A_n(z_t) H_n^{(1)'}(k(\vec{r_t})\rho_t)\cos(n\phi_t), \tag{7}$$

where $H_n^{(1)'}$ is the derivative of the Hankel function with respect to ρ_t expressed as[30]

$$\frac{\partial H_n^{(1)}(k(\vec{r_t})\rho_t)}{\partial \rho_t} = -k(\vec{r_t})H_{n+1}^{(1)}(k(\vec{r_t})\rho_t) + \frac{n}{\rho_t}H_n^{(1)}(k(\vec{r_t})\rho_t). \tag{8}$$

Substituting Eqs. (7) and (8), along with the gradient of the incident field and the waveguide Green's function into Eq. (4), we can calculate the scattered field from the pressure-release cylindrical target at any location in the waveguide. The summation in Eqs. (3) and (7) is truncated at N which is the maximum number of terms sufficient for accurately satisfying the boundary condition. However, since we are also calculating the gradient of the local scattered pressure analytically by Eq. (7), more terms may be needed in order to obtain a stable solution. For targets with small radius such that $ka \ll 1$, we expect only a few terms to be necessary in the approximations.

B. Rigid target

For a target with a rigid boundary, the radial component of the particle velocity must vanish at the surface

$$u(\vec{r_t}) = u_{\text{inc}}(\vec{r_t}) + u_{\text{scat}}(\vec{r_t}) = 0. \tag{9}$$

Note that on the surface of the circular cylinder, the radial component of the particle velocity and the total pressure are related by

$$u(\vec{r_t}) = -\frac{i}{\omega d(\vec{r_t})}\frac{\partial P(\vec{r_t})}{\partial \rho_t}, \tag{10}$$

where d is the medium density. Applying Eqs. (9) and (10) to Green's theorem in Eq. (1), the second term goes to zero and the total scattered field can be calculated from the simplified form below,

$$P_{\text{scat}}(\vec{r}) = -\oint_S [P(\vec{r_t})\nabla_t G(\vec{r}|\vec{r_t})] \cdot \vec{n_t} dA_t. \tag{11}$$

Equations (9) and (10) are used along with Eqs. (7) and (8) to find a least-squares solution for the depth-dependent coefficients $A_n(z_t)$ and the local scattered field. Once these quantities are computed, Eq. (11) is used to calculate the total scattered field at the receiver location.

C. Penetrable target

Let the region outside a homogeneous penetrable cylindrical target be called medium 1 and the region inside the target be called medium 2. The total pressure field in medium 1 is

$$P_1(\vec{r}) = P_{\text{inc}}(\vec{r}) + P_{\text{scat}}(\vec{r}), \tag{12}$$

where $P_{\text{inc}}(\vec{r})$ is calculated using a waveguide acoustic propagation model, as before, and $P_{\text{scat}}(\vec{r})$ is defined in Eq. (3). The pressure field inside the cylindrical target can be expressed as

$$P_2(\vec{r}) \approx \sum_{n=0}^{\infty} B_n(z_t) H_n^{(1)}(k(\vec{r_t})\rho_t)\cos(n\phi_t), \tag{13}$$

where the coefficients $B_n(z_t)$ are determined by application of the boundary conditions.

For this specific case, the two boundary conditions to be satisfied on the surface of the homogeneous penetrable cylindrical target are continuity of acoustic pressure, $P_1(\vec{r_t}) = P_2(\vec{r_t})$, and continuity of the radial component of particle velocity, $u_1(\vec{r_t}) = u_2(\vec{r_t})$. Applying these conditions to Eqs. (3) and (13), we obtain a system of equations that are optimally inverted to find the coefficients $A_n(z_t)$ of Eq. (3) and $B_n(z_t)$ of Eq. (13). Once these coefficients are determined, the total pressure field on the surface of the target can be calculated with Eq. (12) and its gradient with Eqs. (5), (7), and (8). Finally, Green's theorem, Eq. (1), is used to compute the total scattered field level at any location in the waveguide.

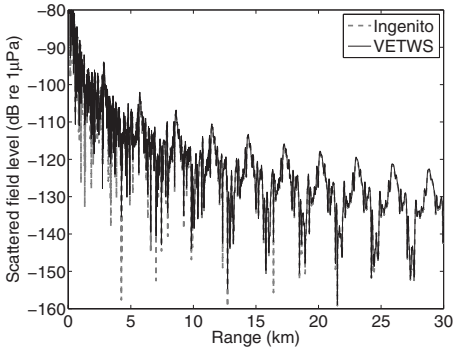


FIG. 2. Scattered field level from the BBN target in a 100 m deep Pekeris waveguide over sandy bottom of Example 1 calculated with the VETWS model and Ingenito model. Source of frequency 415 Hz is co-located with receiver at $z=50$ m; target is centered at $z=50$ m.

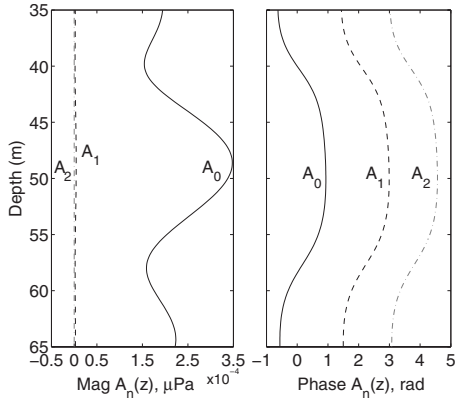


FIG. 3. Plot of magnitude and phase of coefficients $A_n(z_t)$ of Example 1 for target located at 20 km from 415 Hz source.

III. ILLUSTRATIVE EXAMPLES

We implement the scattering model derived in Sec. II for a $L=30$ m long BBN cylindrical target of radius $a=3.5$ cm. The air-filled BBN target satisfies a pressure-release boundary condition so that the equations in Sec. II A are used to compute the scattered field level. A normal mode model[31] is used to calculate the waveguide Green's function for the range-independent cases while the PE model[4] is used for the range-dependent environment. For all calculations, a monostatic configuration is used with a harmonic source of frequency either 415 or 950 Hz co-located with a receiver that measures backscattered fields from the cylindrical target. The acoustic wavelength is much larger than the target's radius, but smaller than its length. The geometry of the problem is shown in Fig. 1. In order to validate our model, we first compare the VETWS model results with Ingenito's scattering solution for the Pekeris waveguide. In addition, we also demonstrate the VETWS result in a layered waveguide with significant sound speed change as a function of target depth and in a range-dependent environment with downslope wedging bottom.

A. Range-independent cases

In order to implement the Ingenito[11,14] scattering model, the following scatter function for a pressure-release cylinder is used:[32,33]

$$S(\alpha,\beta;\alpha_i,\beta_i) = -\frac{kL}{\pi}\mathrm{sinc}\left[\frac{kL}{2}(\cos\alpha_i - \cos\alpha)\right] \times \sum_{m=0}^{\infty} B_m(-j)^m \cos(m[\beta-\beta_i]). \quad (14)$$

The above formula expresses the fact that the cylinder scatters in the vertical like a finite length array through the sinc function[32] and in azimuth through the cylindrical harmonics, where the amplitude of each harmonic for the pressure-release cylinder is given by

$$B_m = -\varepsilon_m j^m \frac{J_m(ka)}{H_m(ka)}. \quad (15)$$

In the above two equations, α and β correspond to the elevation and azimuth angles of the scattered plane waves, α_i and β_i correspond to the elevation and azimuth angles of the incident plane waves, ε_m is the Neumann number defined as $\varepsilon_0=1$, and $\varepsilon_m=2$ for $m\neq 0$, and J_m is the Bessel function of the first kind and order m.

We first compare the VETWS model with the Ingenito model for the BBN target in a 100 m deep Pekeris waveguide over a sandy bottom. The source, receiver, and target centroid are placed at $z=50$ m. The environmental parameters used in this example are water sound speed $c_w=1500$ m/s, bottom sound speed $c_b=1700$ m/s, bottom density $\rho_b=1.9$ g/cm^3, and bottom attenuation $\alpha_b=0.8$ dB/λ.

The total backscattered field level from the target as a function of source-target separation up to 30 km range for a source of frequency 415 Hz is shown in Fig. 2. As expected for this case where the object is contained in the iso-speed water column layer, the scattered levels calculated using the two methods are in very good agreement. Figure 3 shows the depth-dependent coefficients $A_n(z_t)$ of Eqs. (3) and (7) for the Pekeris waveguide problem when the target is located at r

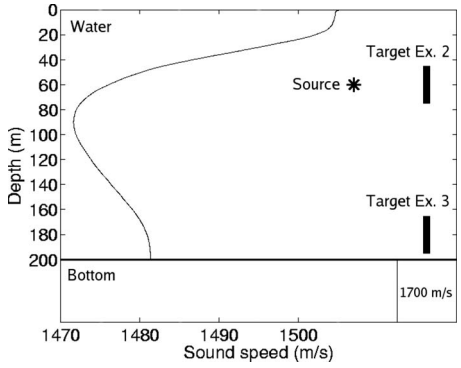


FIG. 4. Mean sound speed profile for Gulf of Maine, during September and October 2006, used in the calculations of the total scattered field for Examples 2 and 3. Source and target locations in depth for both examples are also shown for reference.

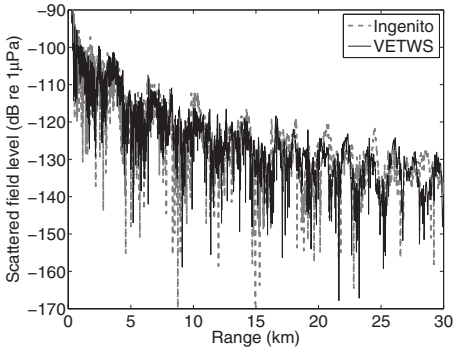


FIG. 5. Scattered field level from the BBN target in the depth-dependent waveguide scenario of Example 2 calculated with the VETWS model and Ingenito model. Source of frequency 415 Hz is co-located with receiver at $z=60$ m; target is centered at $z=60$ m.

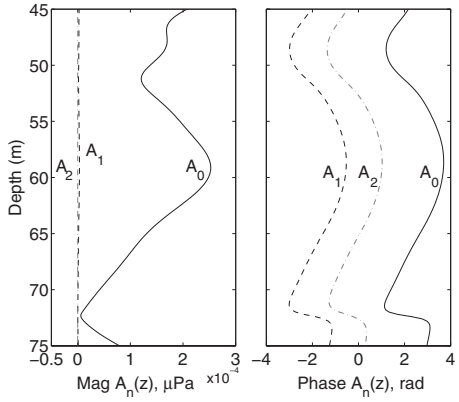


FIG. 6. Plot of magnitude and phase of coefficients $A_n(z_t)$ of Example 2 for target located at 20 km from 415 Hz source.

$=20$ km from the source. Note that only the first three terms in the sum of Eq. (7) are necessary to obtain a stable solution since $ka<0.1$.

Next, we examine the scattering from the BBN target in a 200 m deep waveguide with a depth-dependent sound speed profile over a flat sandy bottom. The sound speed profile used in the calculations, shown in Fig. 4, is the mean sound speed profile derived from 185 expendable bathythermograph measurements acquired during the OAWRS experiment in the Gulf of Maine between September and October 2006. For this environment, we consider two different source-receiver-target configurations illustrated in Fig. 4. In Example 2, the source and receiver are placed at $z=60$ m, and the target is centered at the same depth. In Example 3, the source and receiver are located at $z=60$ m, and the target is centered at $z=180$ m and is suspended 5 m off the bottom. For both examples, calculations are performed for two different frequencies.

In the geometry of Example 2, the target is contained within a layer where the medium sound speed change over the target depth is substantial, about 15 m/s. Figure 5 shows the level of the scattered field calculated using the VETWS model and Ingenito's model for Example 2 for a source of 415 Hz. In this case, the Ingenito model overestimates the scattered field level by roughly 3–5 dB. The modal interference structure of the scattered fields from the two models differ for this case. For the Ingenito model, the interference pattern for the scattered field level depends on the Green's function modal interaction with range at the target center. For the VETWS model, the interference pattern depends on the modal interaction integrated over the depth of the target. Furthermore, the Ingenito model does not account for changes in sound speed along the depth extent of the target. Instead, it uses the sound speed information at the center of the object and assumes that it is contained in an otherwise homogeneous layer. Our VETWS model, on the other hand, takes into account changes in medium sound speed and hence the incident acoustic field refraction along the depth extent of the target. The magnitude and phase of the coefficients $A_n(z_t)$ for Example 2 are shown in Fig. 6 for a target located at $r=20$ km from the 415 Hz source. Because of non-planar, refractive incident field over the target depth, the magnitude and phase of $A_n(z_t)$ in Fig. 6 vary much more drastically compared to Fig. 3 for the iso-speed Pekeris waveguide where these quantities vary smoothly over the target depth. The results for Example 2 using a source of frequency 950 Hz are similar to the lower frequency result, where the Ingenito model overestimates the scattered field level by roughly 3–5 dB.

The results for Example 3, in which the target is centered at $z=180$ m, are shown in Figs. 7(a) and 7(b) for the 415 and 950 Hz sources, respectively. Note that in this ex-

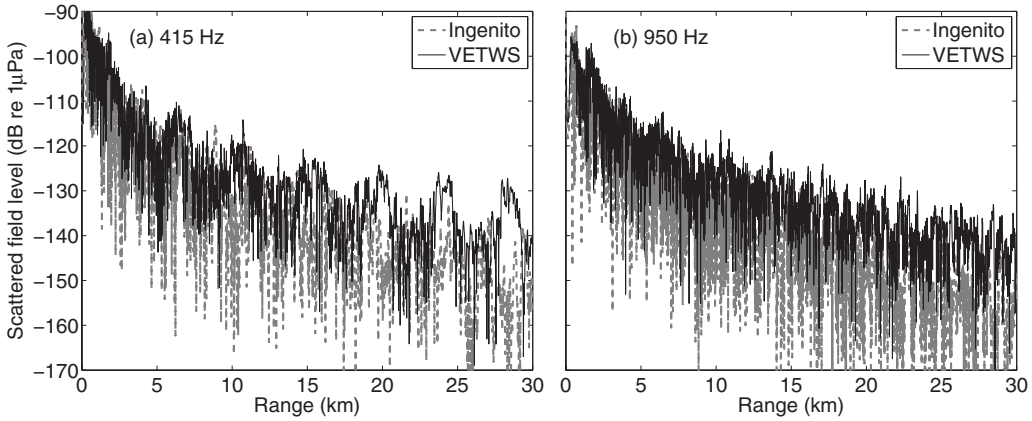


FIG. 7. Scattered field level from the BBN target in the depth-dependent waveguide scenario of Example 3 calculated with the VETWS model and Ingenito model. Source of frequency in (a) 415 Hz and in (b) 950 Hz is co-located with receiver at $z=60$ m; target is centered at $z=180$ m.

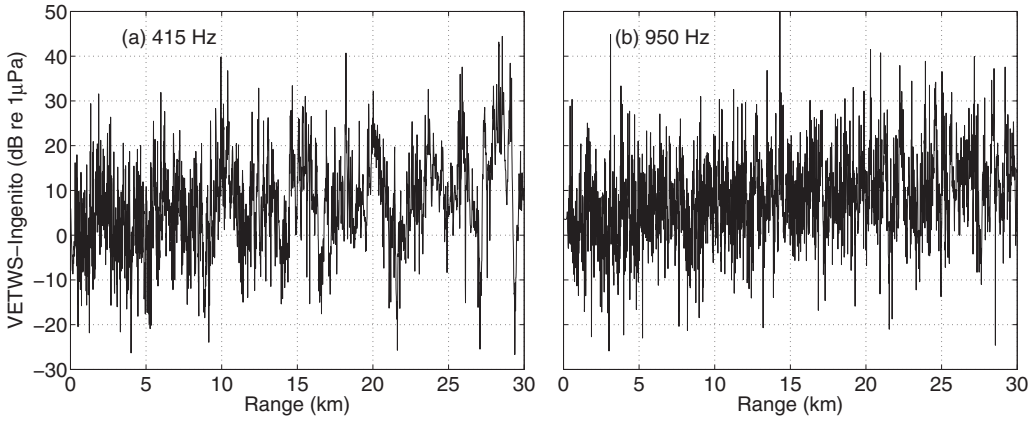


FIG. 8. Difference in scattered field level between the VETWS model and Ingenito model for Example 3 with target centered at $z=180$ m and sources of (a) 415 Hz and (b) 950 Hz.

ample, the target is contained within a layer where the medium sound speed changes by roughly 2 m/s over the target depth. For both frequencies the Ingenito model underestimates the scattered field level by more than 10 dB on average, beyond roughly 20 km source-target separation. The difference in the scattered field level from the VETWS and Ingenito models is shown in Figs. 8(a) and 8(b) for source frequencies of 415 and 950 Hz, respectively. We can observe from these plots that the difference between the two models increases on average with range for this source-receiver-target configuration.

B. Example in a range-dependent environment

In this section, the VETWS model is implemented with the PE to determine the total scattered field from the BBN target in a range-dependent environment. We consider the case of an iso-speed water layer with sound speed $c_w=1500$ m/s, over a downsloping wedge sandy bottom with sound speed $c_b=1700$ m/s. The bathymetry and sound speed used in this example are shown in Fig. 9. The source and receiver are placed at $z=50$ m, and the target is centered at the same depth. Figure 10 shows the total scattered field level for a source of frequency 415 Hz.

For comparison, we also apply the Ingenito model to calculate the total scattered field from the BBN target. Since Ingenito's model is implemented here using a range-independent normal mode model, we approximate the range-dependent waveguide as range-independent using the mean water depth, which in this case is 150 m. Other environmental parameters and the source-receiver-target configuration remain the same. The result is shown in Fig. 10 where we observe that the differences in the total scattered field between the correct range-dependent treatment of the waveguide and its approximation as range-independent are significant. Errors of over 10 dB for ranges greater than 10 km suggest the importance of our approach when calculating the scattered field from vertically cylindrical targets at long ranges from the receiver in real ocean environments.

In the future, the VETWS model developed here will be applied to examine the scattered field data from BBN targets deployed in the Gulf of Maine during the OAWRS experiment with a wide-area sonar in September to October 2006. The current model will be readily applicable to the highly range-dependent environment near Georges Bank where the targets were deployed. A stochastic propagation model will be used in conjunction with the VETWS model to take into account fluctuations in the medium and hence the observed scattered returns from the BBN target.

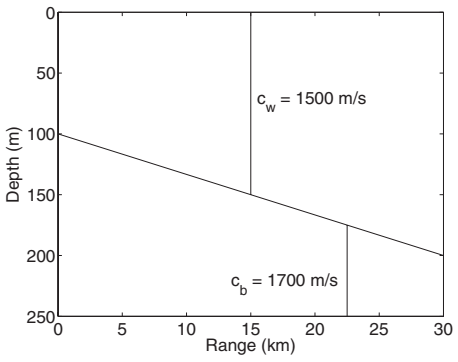


FIG. 9. Bathymetry and sound speed profile used in the calculations of the total scattered field for the range-dependent example.

IV. CONCLUSION

A theoretical model for the scattered field from a penetrable vertically extended cylindrical target in a range-

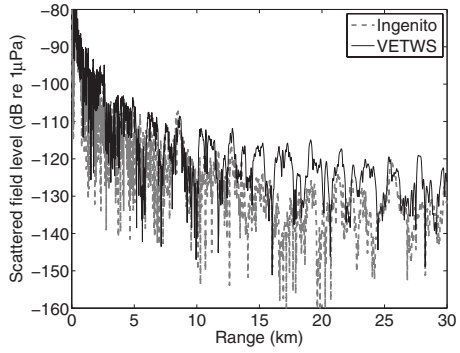


FIG. 10. Scattered field level from the BBN target in the range-dependent environment of Example 4 calculated with the VETWS model. Ingenito model is plotted as reference for a 150 m deep water column. Source of frequency 415 Hz is co-located with receiver at $z=50$ m; target is centered at $z=50$ m.

dependent ocean waveguide is developed from Green's theorem. The model takes into account changes in medium sound speed along the depth extent of the object and hence accounts for non-planar incident field on the target surface. The model is numerically implemented for the BBN target in several different shallow water waveguides. The BBN target is a thin and long air-filled rubber hose often deployed vertically during active acoustic imaging experiments at sea for accurate charting and sonar calibration. The model is validated against the Ingenito waveguide scattering model in an iso-speed Pekeris waveguide where both models lead to identical scattered field levels for the BBN target. In a realistic ocean waveguide, however, with significant change in medium sound speed over the target depth, our analysis shows that the Ingenito model can lead to errors of 10 dB or more in estimating the scattered field level. This is because the Ingenito model assumes that the incident wave over the extent of the target is planar. Our analysis indicates that changes in medium sound speed and hence refraction of the incident acoustic field along the depth extent of a vertically long target must be taken into account in order to accurately estimate its scattered field level in a general ocean waveguide.

ACKNOWLEDGMENTS

This work is funded by the Office of Naval Research, the National Oceanographic Partnership Program, and the Alfred P. Sloan Foundation, with administrative support from Bernard M. Gordon Center for Subsurface Sensing and Imaging Systems. This research is a contribution to the Census of Marine Life.

[1]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1986).
[2]H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography* (Academic, San Diego, 1998).
[3]F. D. Tappert, *The Parabolic Approximation Method*, Lecture Notes in Physics Vol. **70** (Springer, New York, 1977).
[4]M. D. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[5]C. I. Malme, "Development of a high target strength passive acoustic reflector for low-frequency sonar applications," IEEE J. Ocean. Eng. **19**, 438–448 (1994).
[6]W. M. Carey, P. G. Cable, W. L. Siegmann, J. F. Lynch, and I. Rozenfeld, "Measurement of sound transmission and signal gain in the complex Strait of Korea," IEEE J. Ocean. Eng. **27**, 841–852 (2002).
[7]S. Kim, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, G. Edelmann, and T. Akal, "Echo-to-reverberation enhancement using a time reversal mirror," J. Acoust. Soc. Am. **115**, 1525–1531 (2004).
[8]Y. Lai, "Acoustic scattering from stationary and moving targets in shallow water environments with application to humpback whale detection and localization," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA (2004).
[9]P. Ratilal, Y. Lai, D. T. Symonds, L. A. Ruhlmann, J. R. Preston, E. K. Scheer, M. T. Garr, C. Holland, J. A. Goff, and N. C. Makris, "Long range acoustic imaging of the continental shelf environment: The Acoustic Clutter Reconnaissance Experiment 2001," J. Acoust. Soc. Am. **117**, 1977–1998 (2005).
[10]N. C. Makris, P. Ratilal, D. T. Symonds, S. Jagannathan, S. Lee, and R. W. Nero, "Fish population and behavior revealed by instantaneous continental shelf-scale imaging," Science **311**, 660–663 (2006).
[11]F. Ingenito, "Scattering from an object in a stratified medium," J. Acoust. Soc. Am. **82**, 2051–2059 (1987).
[12]J. S. Perkins, W. A. Kuperman, K. D. Heaney, and G. T. Murphy, "Scattering from an object in a three-dimensional ocean," in Proceedings of the 20th Annual International Meeting of the Technical Cooperation Subgroup, Subgroup G, Technical Panel 9 (1991).
[13]A. Sarkissian, "Extraction of a target scattering response from measurements made over long ranges in shallow water," J. Acoust. Soc. Am. **102**, 825–832 (1997).
[14]N. C. Makris and P. Ratilal, "A unified model for reverberation and submerged object scattering in a stratified ocean waveguide," J. Acoust. Soc. Am. **109**, 909–941 (2001).
[15]J. E. Quijano, L. M. Zurk, and D. Rouseff, "Demonstration of the invariance principle for active sonar," J. Acoust. Soc. Am. **123**, 1329–1337 (2008).
[16]R. H. Hackman and G. S. Sammelmann, "Acoustic scattering in an inhomogeneous waveguide: Theory," J. Acoust. Soc. Am. **80**, 1447–1458 (1986).
[17]G. S. Sammelmann and R. H. Hackman, "Acoustic scattering in a homogeneous waveguide," J. Acoust. Soc. Am. **82**, 324–336 (1987).
[18]N. C. Makris, "A spectral approach to 3-D object scattering in layered media applied to scattering from submerged spheres," J. Acoust. Soc. Am. **104**, 2105–2113 (1998).
[19]J. A. Fawcett, "A plane-wave decomposition method for modeling scattering from objects and bathymetry in a waveguide," J. Acoust. Soc. Am. **100**, 183–192 (1996).
[20]J. A. Fawcett, "Modeling acousto-elastic waveguide/object scattering with the Rayleigh hypothesis," J. Acoust. Soc. Am. **106**, 164–168 (1999).
[21]D. K. Dacol and D. G. Roy, "Wave scattering in waveguides," J. Math. Phys. **44**, 2133–2148 (2003).
[22]D. K. Dacol and D. G. Roy, "The partial-wave expansion for scattering in waveguides," J. Acoust. Soc. Am. **120**, 2518–2525 (2006).
[23]L. Cai, D. K. Dacol, D. C. Calvo, and G. J. Orris, "Acoustical scattering by arrays of cylinders in waveguides," J. Acoust. Soc. Am. **122**, 1340–1351 (2007).
[24]G. A. Athanassoulis and A. M. Prospathopoulos, "Three-dimensional acoustic scattering of a source-generated field from a cylindrical island," J. Acoust. Soc. Am. **100**, 206–218 (1996).
[25]G. A. Athanassoulis and A. M. Prospathopoulos, "All-frequency normal-mode solution of the three-dimensional acoustic scattering from a vertical cylinder in a plane-horizontal waveguide," J. Acoust. Soc. Am. **101**, 3371–3384 (1997).
[26]G. A. Athanassoulis and A. M. Prospathopoulos, "Three-dimensional acoustic scattering from a penetrable layered cylindrical obstacle in a horizontal stratified ocean waveguide," J. Acoust. Soc. Am. **107**, 2406–2417 (2000).
[27]R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York, 1983).
[28]P. Ratilal, Y. Lai, and N. C. Makris, "Validity of the sonar equation and Babinet's principle for scattering in a stratified medium," J. Acoust. Soc. Am. **112**, 1797–1816 (2002).
[29]A. Sarkissian, "Multiple scattering effects when scattering from a target in a bounded medium," J. Acoust. Soc. Am. **96**, 3137–3144 (1994).
[30]C. A. Balanis, *Advanced Engineering Electromagnetics* (Wiley, Hoboken, NJ, 1989).
[31]M. Porter and E. L. Reiss, "A numerical method for ocean-acoustic normal-modes," J. Acoust. Soc. Am. **76**, 244–252 (1984).
[32]H. C. van de Hulst, *Light Scattering by Small Particles* (Dover, New York, 1981).
[33]T. K. Stanton, "Sound scattering by cylinders of finite length. I. Fluid cylinders," J. Acoust. Soc. Am. **83**, 55–63 (1988).

A simplified approach to solving finite rough surface scattering problems

J. A. Fawcett
DRDC Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia, B2Y 3Z7, Canada

(Received 5 June 2008; revised 28 January 2009; accepted 28 January 2009)

In this paper a simplified approach is described for the computation of scattering from finite length rough surfaces. In an earlier paper [Fawcett, J. Acoust. Soc. **118**, 1348–1357 (2005)], it was described how to solve such problems by considering a scattering chamber and using the method of wavefield superposition to satisfy the various boundary and continuity conditions. In this paper, a form of the field in the interior of the scattering chamber is assumed and the solution to the scattering problem is reduced to solving a system of equations just along the rough surface. The solution to this system of equations can then be used to compute the scattered or total fields anywhere in the half-space. The results of this approach are compared to those obtained by other methods. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3085645]

PACS number(s): 43.30.Hw, 43.30.Gv [WLS] Pages: 1937–1941

I. INTRODUCTION

In order to compute the field scattered from a rough surface, various authors[1–5] have utilized boundary integral equation methods (BIEMs). In Refs. 1 and 2, exact integral equation solutions to an infinite surface scattering problem are derived with the assumption that the surface is periodic in nature. In Refs. 3 and 4, the integral equations for the unknown fields on an infinite scattering surface are truncated at finite integral limits. These limits can be varied until a stable solution is obtained. In addition, the incident field is tapered in order to minimize the effects of the truncation. In the case that the deformations to the half-space are only inward, one can use the half-space Green's function appropriate to the flat portion of the surface and obtain an integral equation only on the deformed surface.[5] However, this approach is not valid if the deformation contains outward sections. In Ref. 6, an exact spectral formulation is described for the solution of scattering from a finite periodic surface. There are many other numerical approaches (see, for example, Refs. 7 and 8) which can be used to model the scattering from a rough surface.

Wavefield superposition methods have been used by several authors[9–13] in target scattering and range-dependent propagation problems. In Ref. 9 wavefield superposition is used to represent the scattered field external to a target and the computational results are compared to those obtained using a Helmholtz integral representation. In Refs. 10 and 11, wavefield superposition is again used to model target scattering but in the presence of waveguide or half-space boundaries. In Refs. 12 and 13 the method is applied to compute the scattering from range-dependent seabed interfaces. In these cases, the wavefield superposition method involves utilizing multiple sets of point sources to model the fields (possibly elastic) in the seabed and in the water column. In the present paper and in Ref. 14, a novel application of the wavefield superposition method is used to reduce infinite surface (but finite roughness) scattering problems to finite computational domains. In Ref. 14 we described some of the problems associated with modeling scattering from a general finite roughness section, particularly in the case of a distant point source and possibly distant receivers. Truncated integral equation methods, including the extrapolation of the solution using the half-space Green's function in the Helmholtz integral, were discussed. In that reference, the concept of a surrounding scattering chamber (including the rough surface as one of the boundaries) was introduced and wavefield superposition was used to represent the interior and exterior pressure fields. In this approach, there is a set of point sources just exterior to the rough surface, a set of sources just exterior to a bounding arc in the water column, and a set of sources just interior to this same arc. The first two sets of sources are free space sources and represent the interior field in the scattering chamber. The last set of point sources are the appropriate half-space sources which represent the exterior field and ensure that this field satisfies the exterior boundary conditions. The continuity conditions across the arc and the boundary condition along the rough surface allow for the solution of the coefficients of all these sources. In the case that M coefficients are used for each set of sources, then the solution of a system of $3M \times 3M$ equations is required.

In this paper, we start from the same approach as Ref. 14. However, by assuming a representation of the field in the interior, we can reduce the scattering problem to solving a simple $M \times M$ system along the rough boundary. In Sec. II, we describe this new approach and in the numerical examples we examine its accuracy for some rough surface scattering problems using both integral equation approaches and the full method of Ref. 14 for comparison.

II. THEORY

In Fig. 1 a schematic of a pressure-release surface rough surface bounding a constant velocity half-space is shown. The curves C_1 and C_2 comprise the boundary and define the curves for the placement of point sources in the computational method of Ref. 14. The curve C_1 includes the rough surface and small flat sections of $z=0$, the curve C_2 is a

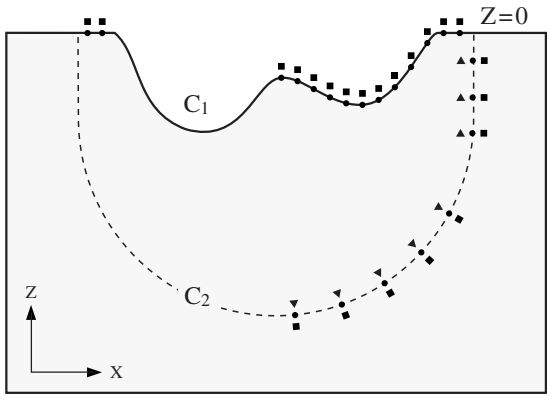


FIG. 1. A schematic diagram showing the bounding curves and point sources which are used in the field representations of Ref. 14 and this paper.

semi-circle in the water half-space. The sources exterior to the scattering chamber (indicated by squares in Fig. 1) represent the field in the interior of the chamber. The sources interior to C_2 (triangles) express the exterior, scattered field and will not be considered in this paper. The positions of the displaced point sources, exterior to C_1 will be denoted as $\boldsymbol{S}_j$, $j=1,\ldots,N$ and those exterior to C_2 as $\boldsymbol{S}_j$, $j=N+1,\ldots,2N$. These sources will have complex amplitudes defined as $\{\hat{a}_j\}$ and $\{b_j\}$, respectively. Thus, we can write for the field interior of $\Omega \equiv C_1 \cup C_2$,

$$p^I(\boldsymbol{R}) = \sum_{C_1} \hat{a}_j G_f(\boldsymbol{R};\boldsymbol{S}_j) + \sum_{C_2} b_j G_f(\boldsymbol{R};\boldsymbol{S}_j), \tag{1}$$

where G_f denotes the free space Green's function,

$$G_f(\boldsymbol{R};\boldsymbol{S}_j) = -i/4H_0^1(k|\boldsymbol{R}-\boldsymbol{S}_j|), \tag{2}$$

$\boldsymbol{R}$ denotes the Cartesian coordinate (x,z) and k denotes the wavenumber in the half-space.

In this paper, instead of using Eq. (1) we will assume a representation of the pressure field in the interior of Ω in the form

$$p^I(\boldsymbol{R}) = \sum_{j=1}^{N} a_j G_f(\boldsymbol{R};\boldsymbol{S}_j) + p_h^{\text{inc}}(\boldsymbol{R}), \tag{3}$$

where p_h^{inc} is the half-space incident field for a pressure-release surface,

$$p_h^{\text{inc}} = -i/4[H_0^1(k\sqrt{(x-x_s)^2+(z-z_s)^2}) - H_0^1(k\sqrt{(x-x_s)^2+(z+z_s)^2})] \tag{4}$$

for a point source at (x_s,z_s). In Eq. (1) there is no explicit representation of the incident field; however, in Ref. 14 it is explicit in the representation of the external field and it implicitly enters into the interior field through the continuity conditions satisfied by the interior and exterior fields along C_2. In Eq. (3) we have reduced the number of point sources to be just those lying to the exterior of C_1 but we have explicitly introduced the half-space incident field. Intuitively, it seems reasonable to consider the pressure field in Ω as composed of the incident field, corresponding to the infinite undeformed surface, and a scattered field represented by point sources along the rough section of the pressure-release surface. The coefficients a_j in Eq. (3) are determined by satisfying the pressure-release boundary condition at the discrete points $\boldsymbol{R}_i$ along C_1,

$$\sum_{j=1}^{N} a_j G_f(\boldsymbol{R}_i;\boldsymbol{S}_j) = -p_h^{\text{inc}}(\boldsymbol{R}_i), \quad i=1,\ldots,N. \tag{5}$$

The expression, Eq. (3), represents not only the interior pressure field but can be used to define its spatial derivatives. Thus, once we have found a_j we can generate the normal derivative of the pressure field along C_1 using the formula

$$\frac{\partial p(\boldsymbol{R}_i)}{\partial n} = \sum_{j=1}^{N} a_j \frac{\partial G_f(\boldsymbol{R}_i;\boldsymbol{S}_j)}{\partial n} + \frac{\partial p_h^{\text{inc}}(\boldsymbol{R}_i)}{\partial n}. \tag{6}$$

Let us consider a spatial point $\boldsymbol{R}=(x,z)$ that lies outside the bounding curve C_2. From Ref. 14, we can compute the pressure field at this point from an integral over just the rough section C_1

$$p(\boldsymbol{R}) = -\int_{C_1} G_h(\boldsymbol{R},\boldsymbol{R}(s')) \frac{\partial p(\boldsymbol{R}(s'))}{\partial n'} ds' + p_h^{\text{inc}}(\boldsymbol{R}). \tag{7}$$

Here $G_h(\boldsymbol{R},\boldsymbol{R}(s'))$ is the half-space Green's function

$$G_h(\boldsymbol{R},\boldsymbol{R}(s')) = -i/4[H_0^1(k\sqrt{(x-x(s'))^2+(z-z(s'))^2}) - H_0^1(k\sqrt{(x-x(s'))^2+(z+z(s'))^2})]. \tag{8}$$

For points $\boldsymbol{R}$ within Ω, the expression of Eq. (3) is used to compute the field. The accuracy of Eq. (7) in computing the pressure field in the domain exterior to Ω is dependent on the accuracy of $\partial p(\boldsymbol{R})/\partial n$ computed using Eq. (6). This, in turn, depends on the accuracy of the representation of the wavefield, Eq. (3) (and its spatial derivatives), near the rough surface C_1.

The method of this paper will also apply for the case of a rigid surface with the appropriate change of boundary conditions and Green's Function. In this case, Eq. (5) becomes

$$\sum_{j=1}^{N} a_j \frac{\partial G_f(\boldsymbol{R}_i;\boldsymbol{S}_j)}{\partial n} = -\frac{\partial p_h^{\text{inc}}(\boldsymbol{R}_i)}{\partial n}, \quad i=1,\ldots,N, \tag{9}$$

where now p_h^{inc} denotes the half-space incident field for a rigid boundary condition along $z=0$. The external field relation, Eq. (7), becomes

$$p(\boldsymbol{R}) = \int_{C_1} \frac{\partial G_h}{\partial n'}(\boldsymbol{R},\boldsymbol{R}(s')) p(s') ds' + p_h^{\text{inc}}(\boldsymbol{R}). \tag{10}$$

In Ref. 14 we used the method of wavefield superposition to solve the continuity and boundary conditions along the boundaries of the numerical scattering chamber. In this paper we have reduced the scattering problem to simply satisfy a boundary condition along C_1. If we allow the superposition point source positions $\boldsymbol{S}_j$ in Eqs. (3) and (9) to approach the boundary, we obtain a boundary integral equation for the unknown source distributions $\alpha(s)$ for these two cases, pressure-release and rigid boundary conditions, respectively,

$$\int_{C_1} G(\boldsymbol{R}(s);\boldsymbol{R}(s'))\alpha(s')ds' = -p_1^{\text{inc}}(\boldsymbol{R}(s)), \tag{11}$$

$$-\alpha(s)/2+\int_{C_1}\frac{\partial G(\boldsymbol{R}(s);\boldsymbol{R}(s'))}{\partial n}\alpha(s')ds'=-\frac{\partial p_2^{\mathrm{inc}}(\boldsymbol{R}(s))}{\partial n}. \tag{12}$$

Here we have used subscripts with the incident fields to emphasize that they are different in the two cases and are the half-space incident fields appropriate to the boundary condition along $z=0$. The advantage of using the boundary integral formulation is that it avoids any questions related to the optimal positioning of the superposition point sources relative to the surface. The disadvantage is that one must account for singularities in the Green's function and its normal derivative as a source point approaches the boundary. In Eq. (12) this results in the term $-\alpha(s)/2$. In Eq. (11) the singularity is logarithmic and is not explicitly shown but can be analytically integrated over a discrete panel. Once the distributions $\alpha(s)$ have been found, then the integral form of Eq. (3) can be used for both the pressure-release and rigid boundary conditions to compute the pressure field within Ω,

$$p^I(\boldsymbol{R})=\int_{c_1}\alpha(s')G_f(\boldsymbol{R};\boldsymbol{R}(s'))ds'+p_h^{\mathrm{inc}}(\boldsymbol{R}). \tag{13}$$

Also once $\alpha(s)$ has been determined, then either Eq. (7) (pressure-release boundary condition) or Eq. (10) (rigid boundary condition) can be used to compute the field exterior to Ω. In order to use Eq. (7), we compute (in a discretized form) the normal derivative of the field along C_1 from

$$\frac{\partial p(\boldsymbol{R}(s))}{\partial n}=\int_{C_1}\frac{\partial G_f(\boldsymbol{R}(s);\boldsymbol{R}(s'))}{\partial n'}\alpha(s')ds'+\frac{\partial p_{\mathrm{inc}}(\boldsymbol{R}(s))}{\partial n} \tag{14}$$

and in order to use Eq. (10) for the rigid boundary condition, we compute the pressure field along C_1 using

$$p(\boldsymbol{R}(s))=\int_{C_1}G_f(\boldsymbol{R}(s);\boldsymbol{R}(s'))\alpha(s')ds'+p^{\mathrm{inc}}(\boldsymbol{R}(s)). \tag{15}$$

In both these cases, since we are computing the field or its normal derivative upon the surface itself, we must use the appropriate analytic evaluation of the singular behavior when the evaluation point $\boldsymbol{R}(s)$ is coincident with one of the discrete points $\boldsymbol{R}(s')$.

In the numerical examples, we will require a method to benchmark the proposed method of this paper. One possible method would be to use the wavefield superposition method of Ref. 14 using all the unknown sources. However, as an independent approach we will implement a standard, truncated BIEM. For the pressure-release and rigid conditions, we will use the following integral formulations:

$$\frac{1}{2}\frac{\partial p(\boldsymbol{R}(s))}{\partial n}+\int_C\frac{\partial G(\boldsymbol{R}(s);\boldsymbol{R}(s'))}{\partial n}\frac{\partial p(s')}{\partial n'}ds'=\frac{\partial p^{\mathrm{inc}}(\boldsymbol{R}(s))}{\partial n}, \tag{16}$$

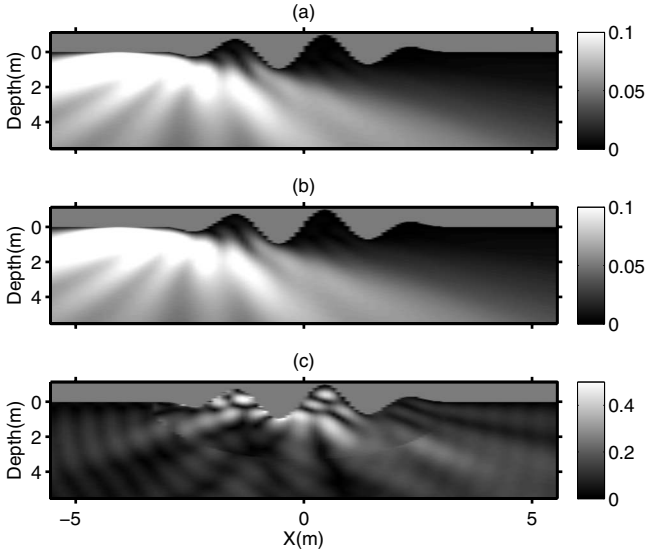


FIG. 2. The total pressure fields (amplitude) for the pressure-release boundary condition: (a) the result from a standard truncated BIEM, (b) the result using the SISC method, and (c) the amplitude of the difference between the two computed fields. The image values in this panel have been scaled so that the maximum value of the grayscale corresponds to 0.5% of the maximum value of the colorbar used in the upper two plots.

$$\frac{p(\boldsymbol{R}(s))}{2}-\int_C\frac{\partial G(\boldsymbol{R}(s);\boldsymbol{R}(s'))}{\partial n'}p(s')ds'=p^{\mathrm{inc}}(\boldsymbol{R}(s)). \tag{17}$$

In this case, the incident field and the Green's function are those for free space, in contrast, to Eqs. (11) and (12) where the half-space Green's functions are used. The integral equations of Eqs. (16) and (17) should be over the entire half-space boundary, but, numerically we will truncate them after a sufficient length (i.e., the curve C will, in general, be longer than C_1) to have achieved a convergent solution.

III. NUMERICAL EXAMPLES

For the numerical examples, we consider a flat surface with a tapered sinusoidal deformation given by

$$z(x)=\sin(\pi x)\cos((x/3)\pi/2),\quad -3<x<3, \tag{18}$$

$$z(x)=0,\quad |x|>3, \tag{19}$$

where increasing depth corresponds to increasingly negative values of z. A 1500-Hz point source is located at a range of 4 m [with respect to (0,0)] and angle of 5° off the x-axis. For the truncated integral equation approach, we consider the computational domain $-15\ \mathrm{m}\leqslant x\leqslant 15\ \mathrm{m}$ with 1500 discrete points used. For the superposition method, we consider the interval $-3.3\ \mathrm{m}\leqslant x\leqslant 3.3\ \mathrm{m}$ to define C_1 of Fig. 1 and the unknown point source distribution using 600 discrete points. The bounding curve C_2 is the semicircle $|\boldsymbol{R}|=3.3$ m. One can certainly obtain accurate solutions using fewer than 600 points but for the error comparisons shown in Figs. 2 and 3 these values were used. Once the point sources have been determined for the interval [−3.3 m 3.3 m], the pressure field can be computed for a much larger numerical grid. Equation (7) [after using Eq. (14) to compute the normal derivative on

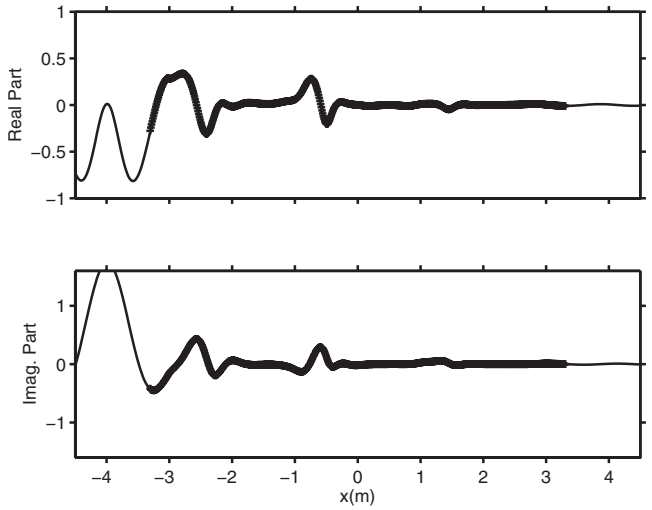


FIG. 3. The solid line shows the real and imaginary parts of the normal derivative of the pressure field along the surface as directly computed by the truncated BIEM. The + signs show the computed values, SISC method, along C_1.

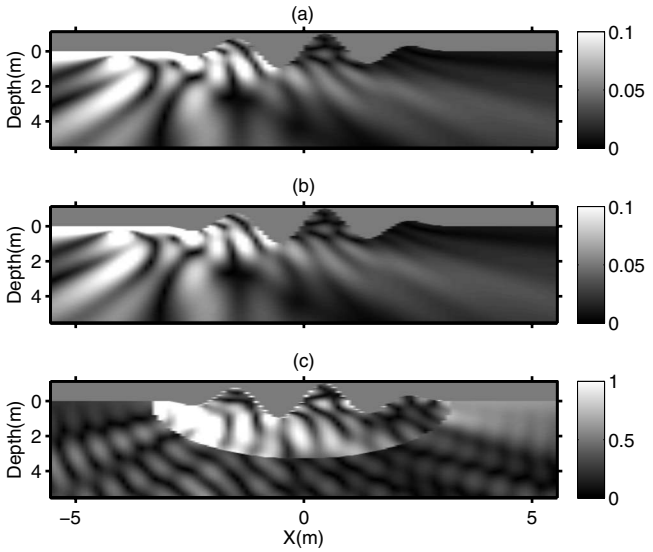


FIG. 4. The total pressure fields (amplitude) for the rigid boundary condition: (a) the result from a standard truncated BIEM, (b) the result using the SISC method, and (c) the amplitude of the difference between the two computed fields. The image values in this panel have been scaled so that the maximum value of the grayscale corresponds to 1% of the maximum value of the colorbar used in the upper two plots.

C_1] is used to compute the pressure field for $|\boldsymbol{R}|>3.3$ m and Eq. (13) is used to compute the pressure field in the interior of Ω. We will refer to the approach of this paper as the surface integral scattering chamber (SISC) method. In Fig. 2 we show the computed total pressure field in the box $[-5.5\ \text{m}\ 5.5\ \text{m}]\times[1.1\ \text{m}\ -5.5\ \text{m}]$ as computed using the standard truncated integral equation approach and the SISC method. In addition, Fig. 2(c) shows the amplitude of the difference between the two solutions scaled by a factor of 1000 so that the maximum amplitude of the colorbar in this plot corresponds to 0.5% the maximum value of the colorbar used in the upper plots. There is a small amount of difference between the BIEM and SISC methods. The differences are largest near the rough scattering surface. There is a slight discontinuity at $|\boldsymbol{R}|=3.3$ m. This is the boundary between the regions where the two different integral representations are used to compute the pressure field using the SISC method. The difference between the two methods is a result of any discretization errors in each of the two methods, BIEM and SISC, but is also a result of any errors in the assumption of Eq. (3). In Fig. 3, we show a comparison of the real and imaginary parts of the normal derivative of the total field as computed by the SISC method using Eq. (14) (+) with those computed using the BIEM method (solid). For the standard BIEM, the normal derivative is solved directly in the interval $[-15\ \text{m}\ 15\ \text{m}]$ using Eq. (16) and the resulting values are shown for a larger interval than is used by the SISC method. As can be seen, the computed values from the two methods show excellent agreement. It is the normal derivative values along C_1 which are used by the SISC method [Eq. (7)] in the computation of the pressure field in the exterior of Ω.

The previous computations are now repeated for a rigid boundary condition using the same numerical parameters as for the pressure-release computations. In Fig. 4 we show the resulting pressure field grids as computed using the BIEM and SISC methods and the difference field, Fig. 4(c). Once again, the differences are small but the error is larger than for the pressure-release boundary condition of Fig. 2 and the discontinuity at the SISC computational boundary $|\boldsymbol{R}|=3.3$ m is evident. In Fig. 5 the computed pressure field upon the scattering surface is shown as computed directly using the standard BIEM and as computed using the solution $\alpha(s)$ from the SISC method [Eq. (15)]. As can be seen, the agreement is excellent.

Finally, we use the SISC method to compute the field scattered by the rigid rough surface for a source 1° off the horizontal at a range of 1000 m. In this case, the truncated BIEM approach cannot be used without using a very large number of points. In order to provide a benchmark for the method of this paper, we use the scattering-chamber/wavefield superposition approach of Ref. 14 (modified for rigid boundary conditions) with 601 free space point source exterior to C_1 $[-3.3\ \text{m}\ 3.3\ \text{m}]$, 601 free space point sources exterior to C_2, and 601 rigid-half-space point sources interior

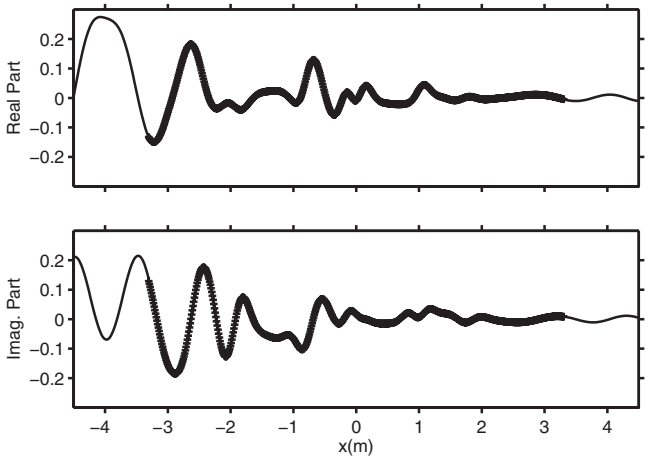


FIG. 5. The solid line shows the real and imaginary parts of the pressure field along the surface as directly computed by the truncated BIEM. The + signs show the computed values, SISC method, along C_1.

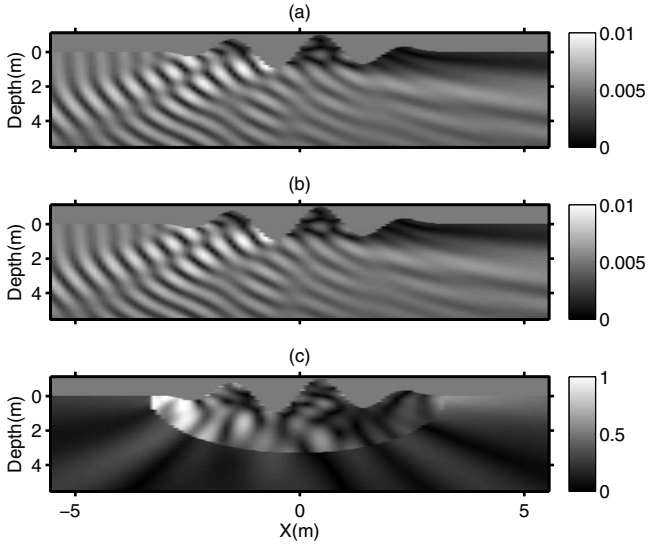


FIG. 6. The total pressure fields for the rigid boundary condition: (a) the result from using the full scattering-chamber approach of Ref. 14, (b) the result using the SISC method, and (c) the amplitude of the difference between the two computed fields. The image values in this panel have been scaled so that the maximum value of the colorbar corresponds to 1% of the maximum value of the grayscale used in the upper two plots. The source is located 1000 m to the left at an angle of 1° with respect to the horizontal.

to C_2. The sources were displaced a distance of 0.04 m from the curves along the normal directions. The resulting total fields are shown for the two methods in Fig. 6. The agreement is excellent. However, the bottom difference plot in Fig. 6 (it is scaled so that the maximum amplitude of the colorbar corresponds to 1% of the maximum amplitude of the colorbar in the upper plots) shows the main differences to be within the SISC interior domain Ω and there are once again discontinuities at the SISC computational boundary $|\boldsymbol{R}|=3.3$ m. The differences in the exterior domain $|\boldsymbol{R}|>3.3$ m are small.

IV. SUMMARY

We have shown in this paper that an infinite surface scattering problem, but with a general finite roughness section, can be accurately solved using a finite interval integral equation. The integral equation was derived by considering a numerical scattering chamber in conjunction with the method of wavefield superposition. Using an assumed representation for the pressure field in the interior of the scattering chamber, the solution to the scattering problem was reduced to solving a single set of equations along the finite rough surface. In the limit of the point sources approaching the scattering surface, integral equations were obtained. For the particular numerical examples considered, the method yielded very accurate results for both pressure-release and rigid boundary conditions.

[1]R. L. Holford, "Scattering of sound waves at a periodic pressure-release surface: An exact solution," J. Acoust. Soc. Am. **70**, 1116–1128 (1981).
[2]D. F. McCammon and S. T. McDaniel, "Application of a new theoretical treatment to an old problem: Sinusoidal pressure release boundary scattering," J. Acoust. Soc. Am. **78**, 149–156 (1985).
[3]E. I. Thorsos, "The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum," J. Acoust. Soc. Am. **83**, 78–92 (1988).
[4]R. T. Marchand and G. S. Brown, "On the use of finite surfaces in the numerical prediction of rough surface scattering," IEEE Trans. Antennas Propag. **47**, 600–604 (1999).
[5]T. W. Dawson and J. A. Fawcett, "A boundary integral equation method for acoustic scattering in a waveguide with nonplanar surfaces," J. Acoust. Soc. Am. **87**, 1110–1125 (1990).
[6]A. Kashihara and J. Nakayama, "Scattering of TM plane wave from a finite periodic surface," Electron. Commun. Jpn., Part 2: Electron. **89**, 10–19 (2006).
[7]D. A. Miles, R. N. Hewitt, M. K. Donnelly, and T. Clarke, "Forward scattering of pulses from a rough sea surface by Fourier synthesis of parabolic equation solutions," J. Acoust. Soc. Am. **114**, 1266–1280 (2003).
[8]R. A. Stephen, "Modeling sea surface scattering by the time-domain finite-difference method," J. Acoust. Soc. Am. **100**, 2070–2078 (1996).
[9]R. D. Miller, E. T. Moyer, Jr., H. Huang, and H. Uberall, "A comparison between the boundary element method and the wave superposition approach for the analysis of the scattered fields from rigid bodies and elastic shells," J. Acoust. Soc. Am. **89**, pp. 2185–2196 (1991).
[10]A. Sarkissian, "Method of superposition applied to scattering from a target in shallow water," J. Acoust. Soc. Am. **95**, 2340–2345 (1994).
[11]I. Lucifredi and H. Schmidt, "Subcritical scattering from buried elastic shells," J. Acoust. Soc. Am. **120**, 3566–3583 (2006).
[12]J. E. Murphy, G. Li, S. Chin-Bing, and D. King, "Multifilament source model for short-range underwater acoustic problems involving penetrable ocean bottoms," J. Acoust. Soc. Am. **99**, 845–850 (1996).
[13]A. T. Abawi and M. B. Porter, "Propagation in an elastic wedge using the virtual source technique," J. Acoust. Soc. Am. **121**, 1374–1382 (2007).
[14]J. A. Fawcett, "A scattering-chamber approach for solving finite rough surface scattering problems," J. Acoust. Soc. Am. **118**, 1348–1357 (2005).

Finite-difference modeling of the monopole acoustic logs in a horizontally stratified porous formation

Wei Guan,[a] Hengshan Hu, and Xiao He
Department of Astronautics and Mechanics, Harbin Institute of Technology, P.O. Box 344, 92 West Dazhi Street, Harbin 150001, China

(Received 22 April 2008; revised 19 January 2009; accepted 22 January 2009)

Monopole acoustic logs in a homogeneous fluid-saturated porous formation can be simulated by the real-axis integration (RAI) method to analytically solve Biot's equations [(1956a) J. Acoust. Soc. Am. **28**, 168–178; (1956b) J. Acoust. Soc. Am. **28**, 179–191; (1962) J. Appl. Phys. **33**, 1482–1498], which govern the wave propagation in poro-elastic media. Such analytical solution generally is impossible for horizontally stratified formations which are common in reality. In this paper, a velocity-stress finite-difference time-domain (FDTD) algorithm is proposed to solve the problem. This algorithm considers both the low-frequency viscous force and the high-frequency inertial force in poro-elastic media, extending its application to a wider frequency range compared to existing algorithms which are only valid in the low-frequency limit. The perfectly matched layer (PML) is applied as an absorbing boundary condition to truncate the computational region. A PML technique without splitting the fields is extended to the poro-elastic wave problem. The FDTD algorithm is validated by comparisons against the RAI method in a variety of formations with different velocities and permeabilities. The acoustic logs in a horizontally stratified porous formation are simulated with the proposed FDTD algorithm.
 [DOI: 10.1121/1.3081518]

PACS number(s): 43.30.Ky, 43.20.Gp, 43.20.Bi, 43.40.Ph [RAS] Pages: 1942–1950

I. INTRODUCTION

Acoustic logging is a useful technique to explore formation properties in geophysics and oil engineering. Important applications of acoustic logging include evaluation of porous formation parameters such as porosity and permeability (Williams *et al.*, 1984; Tang and Cheng, 1996), detection of fractures and measurement of fracture permeability (Hardin *et al.*, 1987; Hornby *et al.*, 1989), and imaging structural characteristics of formation away from the borehole (Hornby, 1989; Tang *et al.*, 2007).

In order to study the relations between acoustic logging responses and properties of porous formations, Rosenbaum (1974) first simulated the axisymmetric responses of the monopole acoustic logs in a homogeneous fluid-saturated porous formation using the real-axis integration (RAI) method to solve Biot's equations (Biot, 1956a, 1956b, 1962), which governs the acoustic wave propagation in poro-elastic media. Then Schmitt *et al.* (1988) analyzed the effects of the permeability, the pore geometry, the properties of the pore fluid, and the borehole radius on the acoustic logs. With the RAI method, Schmitt (1988) simulated the monopole acoustic logs in radially layered fluid-saturated porous formations. Unfortunately, this RAI approach, being analytical, does not work for the common case where the porous formation is horizontally stratified. To solve this, one seeks instead a numerical approach, such as finite-difference, finite-element, or boundary-element method. In this work, a finite-difference time-domain (FDTD) scheme is developed.

There have been several FDTD algorithms presented in the literature for the modeling of acoustic wave propagation in porous media (e.g., Zhu and McMechan, 1991; Dai *et al.*, 1995; Dong *et al.*, 2000; Zeng *et al.*, 2001; Zeng and Liu, 2001). These algorithms are valid and correct only in the low-frequency limit, where the fluid flow in pore space is practically a Poiseuille flow (Biot, 1956a) and the viscous force term in Biot's equations can be treated as independent of frequency. This is so since at low-frequency, the inertial force is much less than the viscous force and hence can be safely ignored. Recently, improved algorithms considering the inertial effect in the viscous fluid flow in poro-elastic media are proposed by Masson *et al.* (2006) and Guan and Hu (2008), but still not working very well for inertial flow in the high-frequency range, where the inertial force dominates. Both works use Taylor expansion to approximate the frequency correction function for the viscous force in the time domain. This idea is extended in the current work to model the acoustic logs in a horizontally stratified fluid-saturated porous formation.

In numerical simulations of wave propagation in an unbounded medium, some absorbing boundary condition needs to be employed to eliminate the spurious reflections from the artificial boundaries of the computational region. In the past 30 years, a great many absorbing boundary conditions have been developed: for example, one-wave equations (Clayton and Engquist, 1977; Reynolds, 1978; Liao *et al.*, 1984; Higdon, 1991), damping layers or sponge zones (Cerjan *et al.*, 1985; Kosloff and Kosloff, 1986; Sochacki *et al.*,1987), and optimal boundary conditions (Peng and Toksöz, 1995). These absorbing boundary conditions, although successful in many applications, are only effective for waves within a specific

[a]Electronic mail: gwllzh@tom.com

range of incidence angles and frequencies. Berenger (1994) proposed a highly effective perfectly matched layer (PML) as an absorbing boundary condition for electromagnetic waves. The PML has the remarkable property of having a zero reflection coefficient for all incidence angles and frequencies before discretization. Chew and Liu (1996) and Hastings *et al.* (1996) independently demonstrated that the PML can also be applied for acoustic waves. These early schemes require field-splitting in the PML to avoid convolutional operations in the time domain, which means the equations in the PML are different from those in the computational region and computational complexity is increased. Roden and Gedney (2000) proposed a nonsplitting PML technique for finite-difference modeling of electromagnetic wave problems, featuring a recursive convolution evaluation and identical equation form in both PML and the computational region. A similar scheme was used by Wang and Tang (2003) for finite-difference modeling of elastic wave propagation. In this paper, the nonsplitting PML schemes is extended to solve the poro-elastic wave problems.

In Secs. II and III, after a review of Biot's theory of wave propagation in poro-elastic media, we adopt the method of complex stretched coordinates (Chew and Weedon, 1994; Liu and Tao, 1997; Wang and Tang, 2003) to formulate the FDTD discretization of Biot's equations. Then we show the comparisons between the FDTD and the RAI methods of the simulated waveforms in variety of formations with different velocities and permeabilities to validate the FDTD algorithm. Finally, we give numerical simulation examples of the acoustic logs in a horizontally stratified porous formation.

II. FORMULATION

A. Biot's equations

According to Biot (1956a, 1956b, 1962), assuming an $e^{-i\omega t}$ time dependence of all field quantities, the equations that describe the acoustic field in an isotropic, homogeneous fluid-saturated porous medium can be expressed as follows:

$$-i\omega\mathbf{w} = (-\nabla p + \omega^2\rho_f\mathbf{u})\kappa/\eta, \tag{1}$$

$$\nabla\cdot\boldsymbol{\tau} = -\omega^2(\rho\mathbf{u} + \rho_f\mathbf{w}), \tag{2}$$

$$\boldsymbol{\tau} = (H-2G)(\nabla\cdot\mathbf{u})\mathbf{I} + C(\nabla\cdot\mathbf{w})\mathbf{I} + G(\nabla\mathbf{u} + \nabla\mathbf{u}^{\mathbf{T}}), \tag{3}$$

$$-p = C\nabla\cdot\mathbf{u} + M\nabla\cdot\mathbf{w}, \tag{4}$$

where $\mathbf{u}$ is the displacement of the solid phase, $\mathbf{w}$ is the relative flow between the fluid and the solid phase, $\boldsymbol{\tau}$ is the bulk stress tensor, $\mathbf{I}$ is the identity tensor, p is the pore fluid pressure, ρ_f and η are the density and viscosity of the pore fluid, respectively, ρ is the density of the formation, G is the shear modulus of the formation, H, C, and M are porous medium moduli as defined by Biot (1962). Note that κ is the dynamic permeability defined by Johnson *et al.* (1987) with the following form:

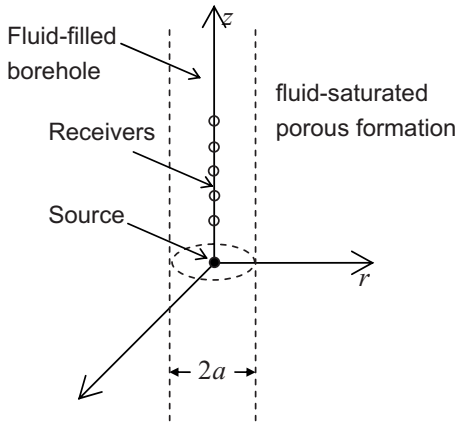


FIG. 1. Schematic of the acoustic logging model. The fluid-filled borehole of radius a is embedded in a horizontally stratified fluid-saturated porous formation. The excited source and the receivers are along the borehole axis, and the source is located at the origin.

$$\kappa(\omega) = \kappa_0\left[\left(1-\frac{4i\omega}{m\omega_c}\right)^{1/2} - i\frac{\omega}{\omega_c}\right]^{-1}, \tag{5}$$

where κ_0 is the Darcy permeability. Note the term $(1-4i\omega/m\omega_c)^{1/2}$ denotes the frequency correction function for the viscous force term in Biot's equations. Here $\omega_c = \phi\eta/\alpha_\infty\rho_f\kappa_0$ is a critical frequency, where α_∞ is the tortuosity. And m is a dimensionless parameter given by $m = \phi\Lambda^2/\alpha_\infty\kappa_0$, where Λ is the weighted volume-to-surface radio. For typical sandstones, $m \approx 8$.

In the low-frequency limit ($\omega \ll \omega_c$), the viscous force dominates and the fluid flow in pore space controlled by viscous shearing is practically a Poiseuille flow. The inertial force increases gradually with frequency and matches the viscous force at $\omega = \omega_c$. Beyond that, the inertial force starts to dominate and the viscous boundary layers first develop. These viscous boundary layers shrink with increasing frequency and at $\omega \gg \omega_c$, the fluid flow in pore space switches to a high-frequency inertial flow. The algorithm to be developed in this work does not drop the inertial term in the viscous fluid flow, thus can be applied at higher frequencies. Recall that FDTD algorithms for the acoustic logs simulation in literature are valid only in the low-frequency limit ($\omega \ll \omega_c$).

B. Finite-difference implementation

A schematic of the acoustic logging model is shown in Fig. 1, where a fluid-filled borehole is embedded in an unbounded fluid-saturated porous formation. With a cylindrical coordinate system (r, z, θ), the borehole axis lies along the central axis z of the cylinder, and the monopole point acoustic source is located at the origin. All field quantities are independent of θ due to axisymmetry.

For the FDTD discretization of Biot's equations governing the acoustic field in the porous formation, we need to rewrite Eqs. (1)–(4) in terms of velocities and stresses in the cylindrical coordinates. Substituting Eq. (5) into Eq. (1), and setting $D_1 = 4/m\omega_c$ and $D_2 = 1/\omega_c$, we get

$$\frac{\eta}{\kappa_0}(\sqrt{1-i\omega D_1}-i\omega D_2)\begin{bmatrix} v_{wr} \\ v_{wz} \end{bmatrix} = -\begin{bmatrix} \frac{\partial p}{\partial r} \\ \frac{\partial p}{\partial z} \end{bmatrix} + i\omega\rho_f \begin{bmatrix} v_{ur} \\ v_{uz} \end{bmatrix}, \tag{6}$$

where $v_{wq}=-i\omega w_q$ and $v_{uq}=-i\omega u_q$ $(q=r,z)$ are the velocity components of the relative flow and the solid phase motion in the i-direction, respectively. Equation (2) can be written as

$$-i\omega\rho\begin{bmatrix} v_{ur} \\ v_{uz} \end{bmatrix} - i\omega\rho_f\begin{bmatrix} v_{wr} \\ v_{wz} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial r}+\frac{1}{r} & -\frac{1}{r} & 0 & \frac{\partial}{\partial z} \\ 0 & 0 & \frac{\partial}{\partial z} & \frac{\partial}{\partial r}+\frac{1}{r} \end{bmatrix}\begin{bmatrix} \tau_{rr} \\ \tau_{\theta\theta} \\ \tau_{zz} \\ \tau_{rz} \end{bmatrix}, \tag{7}$$

where τ_{rr}, $\tau_{\theta\theta}$, τ_{zz}, and τ_{rz} are the components of the bulk stress tensor $\boldsymbol{\tau}$. Similarly, the rest Biot's equations can be expressed in terms of velocities and stresses in the cylindrical coordinates.

In this work, the PML is used as the absorbing boundary condition. To introduce the PML, we will modify the equations in the complex stretched coordinates. In the frequency domain, the complex coordinate-stretching variable is chosen as $s_q=1+\Omega_q/i\omega$ $(q=r,z)$, where Ω_q is the stretching function with respect to q. In the PML, the regular coordinate variable q is replaced by the complex coordinate variable $\tilde{q}$, and the spatial derivative $\partial/\partial\tilde{q}$ is expressed in terms of the complex coordinate-stretching variable as $\partial/\partial\tilde{q}=(1/s_q)\partial/\partial q$. Thus, the r components of Eqs. (6) and (7), for example, are changed to

$$\frac{\eta}{\kappa_0}(\sqrt{1-i\omega D_1}-i\omega D_2)v_{wr}-i\omega\rho_f v_{ur} = -\frac{1}{s_r}\frac{\partial p}{\partial r}, \tag{8}$$

$$-i\omega(\rho v_{ur}+\rho_f v_{wr}) = \left(\frac{1}{s_r}\frac{\partial}{\partial r}+\frac{1}{\tilde{r}}\right)\tau_{rr}-\frac{\tau_{\theta\theta}}{\tilde{r}}+\frac{1}{s_z}\frac{\partial\tau_{rz}}{\partial z}, \tag{9}$$

where $\tilde{r}=\int_0^r s_r(r')dr'=r(1+\bar{\Omega}_r/i\omega)$ and $\bar{\Omega}_r=1/r\int_0^r\Omega_r(r')dr'$.

Before discretization, we transform the equations from the frequency domain to the time domain in a way similar to Wang and Tang (2003). For example, in the time domain, Eqs. (8) and (9) become

$$\frac{\eta}{\kappa_0}\left(\sqrt{1+D_1\frac{\partial}{\partial t}}+D_2\frac{\partial}{\partial t}\right)v_{wr}+\rho_f\frac{\partial v_{ur}}{\partial t} = -(1+\phi_r)\frac{\partial p}{\partial r}, \tag{10}$$

$$\rho\frac{\partial v_{ur}}{\partial t}+\rho_f\frac{\partial v_{wr}}{\partial t} = \left[(1+\phi_r)\frac{\partial}{\partial r}+\left(\frac{1+\bar{\phi}_r}{r}\right)\right]\tau_{rr} - \left(\frac{1+\bar{\phi}_r}{r}\right)\tau_{\theta\theta}+(1+\phi_z)\frac{\partial\tau_{rz}}{\partial z}, \tag{11}$$

where $\phi_q=-\Omega_q e^{-\Omega_q t}*$, $(q=r,z)$, and $\bar{\phi}_r=-\bar{\Omega}_r e^{-\bar{\Omega}_r t}*$ are convolution operators introduced by Wang and Tang (2003). Equations (10) and (11) are modified Biot's equations in the PML region. These equations automatically reduce to Biot's equations in the computational region, where Ω_q and $\bar{\Omega}_r$ become zero.

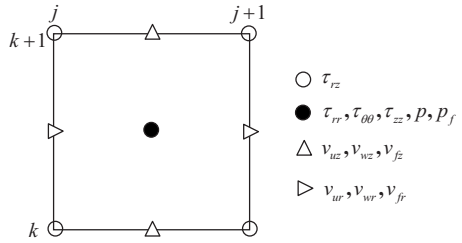


FIG. 2. The velocity-stress FDTD staggered grid in the cylindrical coordinates.

Now we discretize the velocities and stresses by using a staggered finite-difference grid (e.g., Virieux, 1986), as shown in Fig. 2. The grid sizes Δr and Δz in the r- and z-directions, respectively, are constant. To discretize $\sqrt{1+D_1\partial/\partial t}$ $(D_1=4/m\omega_c)$ in Eq. (10), a Taylor expansion in the frequency domain under the condition that $|4\omega/m\omega_c|<1$ gives

$$\sqrt{1-\frac{4i\omega}{m\omega_c}} = 1-\frac{2i\omega}{m\omega_c}+\frac{2}{m^2}\left(\frac{\omega}{\omega_c}\right)^2+o\left(\frac{\omega^3}{\omega_c^3}\right). \tag{12}$$

For a typical sandstone (where $m\approx 8$) and $\omega<\omega_c$, the third term in the right hand side of Eq. (12) is higher order and can be ignored. Hence in the time domain the term $\sqrt{1+D_1\partial/\partial t}$ can be approximated as $1+(D_1/2)\partial/\partial t$.

Then, applying the second-order central difference approximations, one obtains

$$\frac{\eta}{\kappa_0}\left[\frac{1}{2}+\left(\frac{D_1}{2}+D_2\right)\frac{1}{\Delta t}\right]v^{n+1/2}_{wr(j,k+1/2)}+\frac{\rho_f}{\Delta t}v^{n+1/2}_{ur(j,k+1/2)} = \frac{\eta}{\kappa_0}\left[-\frac{1}{2}+\left(\frac{D_1}{2}+D_2\right)\frac{1}{\Delta t}\right]v^{n-1/2}_{wr(j,k+1/2)}+\frac{\rho_f}{\Delta t}v^{n-1/2}_{ur(j,k+1/2)} - \left(\frac{\partial p^n}{\partial r}+P^n_{pr}\right)_{(j,k+1/2)}, \tag{13}$$

where n and $(j,k+1/2)$ are the indices of time and space steps, respectively, Δt is the time step, and P^n_{pr} denotes convolution integral,

$$P^n_{pr} = -\Omega_r\int_0^{n\Delta t} e^{-\Omega_r(n\Delta t-t)}\frac{\partial p(r,t)}{\partial r}dt. \tag{14}$$

The above integral can be further approximated with the trapezoidal rule, where

$$P^n_{pr} = e^{-\Omega_r\Delta t}P^{n-1}_{pr}-\frac{1}{2}\Omega_r\Delta t\left(e^{-\Omega_r\Delta t}\frac{\partial p^{n-1}}{\partial r}+\frac{\partial p^n}{\partial r}\right). \tag{15}$$

Similar to the discretization of Eq. (10), Eq. (11) can be discretized as

$$\rho v_{ur(j,k+1/2)}^{n+1/2} + \rho_f v_{wr(j,k+1/2)}^{n+1/2}$$
$$= \rho v_{ur(j,k+1/2)}^{n-1/2} + \rho_f v_{wr(j,k+1/2)}^{n-1/2} + \Delta t\left[\left(\frac{\partial \tau_{rr}^n}{\partial r} + P_{rrr}^n\right) + \left(\frac{\tau_{rr}^n - \tau_{\theta\theta}^n + Q_{r\_\theta r}^n}{r}\right) + \left(\frac{\partial \tau_{rz}^n}{\partial z} + P_{rzz}^n\right)\right]_{(j,k+1/2)}, \quad (16)$$

where

$$P_{rrr}^n = e^{-\Omega_r \Delta t} P_{rrr}^{n-1} - \frac{1}{2}\Omega_r \Delta t\left(e^{-\Omega_r \Delta t}\frac{\partial \tau_{rr}^{n-1}}{\partial r} + \frac{\partial \tau_{rr}^n}{\partial r}\right), \quad (17)$$

$$Q_{r\_\theta r}^n = e^{-\bar{\Omega}_r \Delta t} Q_{r\_\theta r}^{n-1} - \frac{1}{2}\bar{\Omega}_r \Delta t[e^{-\bar{\Omega}_r \Delta t}(\tau_{rr}^{n-1} - \tau_{\theta\theta}^{n-1}) + \tau_{rr}^n - \tau_{\theta\theta}^n], \quad (18)$$

$$P_{rzz}^n = e^{-\Omega_z \Delta t} P_{rzz}^{n-1} - \frac{1}{2}\Omega_z \Delta t\left(e^{-\Omega_z \Delta t}\frac{\partial \tau_{rz}^{n-1}}{\partial z} + \frac{\partial \tau_{rz}^n}{\partial z}\right). \quad (19)$$

With Eqs. (13) and (16), $v_{ur(j,k+1/2)}^{n+1/2}$ and $v_{wr(j,k+1/2)}^{n+1/2}$ can be solved. The other velocity and stress field quantities in the porous formation can also be obtained from Biot's equations with a similar discretization.

The acoustic field in the fluid-filled borehole is governed by the velocity-pressure fluid equations. The discretization of the field quantities in the borehole and the treatment of the point source on the borehole axis have been given by Guan and Hu (2008) and are incorporated in Appendix A.

It is worth noting that the field quantities at medium interfaces require special treatment. Two kinds of interfaces are involved in this work. One is the interface between two layers in the porous formation, and the other is the borehole wall between the borehole fluid and the porous formation. All the interfaces are considered as open pore interface, which permits relative fluid flow across it. The discretization of the field quantities at the interface between two layers has an identical form with those in homogeneous porous medium, with the only difference that the effective medium parameter (average of the two, see e.g., Taflove and Hagness, 2005) should be used. Note, however, that effective medium parameter cannot be applied in the case of borehole wall, since the governing equations in the borehole are different from those in the formation. A solution to this issue using the boundary conditions of the open pore interface is given in Appendix B.

III. NUMERICAL SIMULATIONS

For all simulations of acoustic logs in the following work, the borehole radius is set to 0.10 m and the borehole fluid velocity and density are assumed to be 1500 m/s and 1000 kg/m^3, respectively. A stretching function of $\Omega_q(q) = -V_{\max}(a\gamma + b\gamma^2)\ln \alpha / T$, $(q = r, z)$ is taken, as given in Wang and Tang (2003), where $V_{\max}$ is the maximum wave velocity, $\gamma = q/T$, T is the width of the PML, $a = 0.25$, $b = 0.75$, and $\alpha = 10^{-6}$ denotes a predefined level of wave absorption.

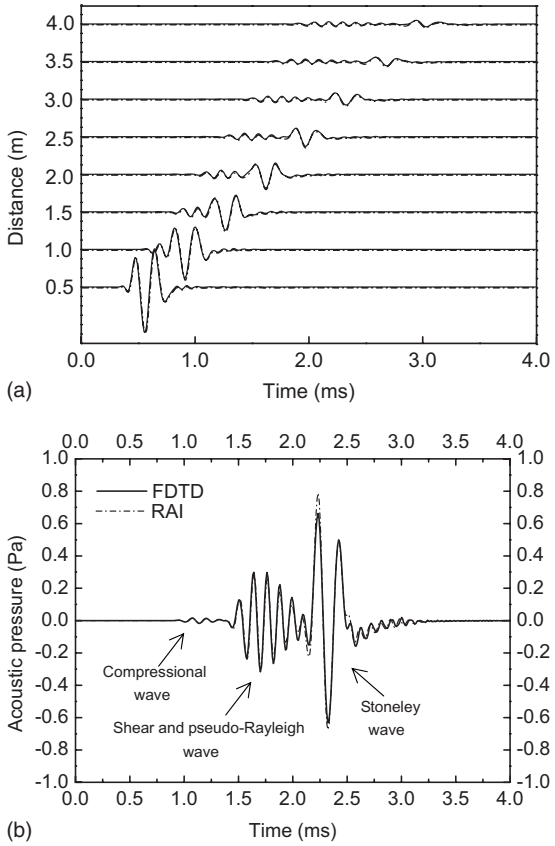


FIG. 3. Comparisons between the FDTD and the RAI methods of the simulated waveforms of the acoustic logs in a homogeneous porous formation (Formation I). The center frequency of the source is $f_0 = 6$ kHz. (a) Pressure waveforms at different locations along the borehole axis. (b) The waveforms of (a) at the location $z = 3.0$ m.

A. Homogeneous porous formations

To check the FDTD algorithm, comparisons of the FDTD against the RAI methods of the simulated acoustic logs for the case with the homogeneous porous formations around the borehole are performed.

In Fig. 3(a) the simulated pressure waveforms at different locations along the borehole axis obtained from the FDTD algorithm are compared with those from the RAI method. The waveforms are normalized with respect to the peak value of the response at a location of receiver-to-source distance $z = 0.5$ m. The formation parameters are listed in the first column (Formation I) of Table I. The formation has a compressional velocity of 3975 m/s and a shear velocity of 2455 m/s. The source has a center frequency and a pulse width of $f_0 = 6$ kHz and $T_c = 0.5$ ms, respectively. The finite-difference grid sizes are set to $\Delta r = \Delta z = 0.0125$ m according to the formula $\Delta r = V_{\min}/10 f_{\max}$ in Alford *et al.* (1974), where $V_{\min}$ is the minimum wave velocity, and $f_{\max}$ is the maximum source frequency where $f_{\max} = 2f_0$. The time step is chosen as $\Delta t = 2 \times 10^{-3}$ ms, following the Courant stability condition $\Delta t < \Delta r / \sqrt{2} V_{\max}$. This condition, which was proposed for FDTD modeling of electromagnetic waves, has been demonstrated by Masson *et al.* (2006) to be valid for poro-elastic waves. There are 160 and 320 cells in the r- and z-directions of the computational region, respectively. The PML contains 40 cells in all directions outside the computa-

TABLE I. Parameters of the porous formations. The relationships between ϕ and K_b and G_b are assumed to obey the experimental results of Vernik (1994).

Parameter	Property	Formation I	Formation II	Formation III
ϕ	Porosity (%)	20	10	40
κ_0	Darcy permeability (Darcy)	1	0.01	10
K_b	Frame bulk modulus (GPa)	14.39	22.63	10.93
G_b	Frame shear modulus (GPa)	13.99	22.23	3.01
K_s	Solid bulk modulus (GPa)	35.70	35.70	33.17
K_f	Pore fluid bulk modulus (GPa)	2.25	2.25	2.0
ρ_s	Solid density (kg/m^3)	2650	2650	2670
ρ_f	Pore fluid density (kg/m^3)	1000	1000	1290
η	Pore fluid viscosity (Pa s)	10^{-3}	10^{-3}	10^{-4}
α_∞	Tortuosity	3	3	3
m	Dimensionless parameter	8	8	8

tional region. The waveforms at different locations from $z=0.5$ m to $z=4.0$ m with the two methods show excellent agreement, as depicted in Fig. 3(a). More details at $z=3$ m are given in Fig. 3(b). It is clear that the waveforms consist of three different wave groups: a compressional, a shear, and pseudo-Rayleigh, and a Stoneley wave group. For the two methods, Fig. 3(b) shows that the phase and the amplitude of the three wave groups agree well, with only a small deviation on the Stoneley wave amplitude.

Analysis shows that the above deviation is caused by the discrete Fourier transforms (DFTs) in the RAI method and the finite-difference discretization in the FDTD method. Two DFTs are required in RAI to transform the field quantities from frequency-wavenumber domain to time-space domain. But the choice of DFT parameters turns out to have an influence on the accuracy of the RAI simulated waveforms. This also explains the obvious deviations of the simulated waveforms between the FDTD and the RAI methods in Guan and Hu (2008). In Fig. 3, the sampling intervals and the sampling numbers of the two DFTs in the RAI method are set to $D_z=4\times10^{-3}$ m, $N_{kz}=10240$, $D_f\approx10$ Hz, and $N_{ft}=10240$. Note that compared to those in Guan and Hu (2008), here D_z and D_f are reduced by ten times and N_{kz} and N_{ft} are increased by ten times, resulting more accurate results. Recall that D_z and D_f are the sampling intervals in the space and the frequency domains, respectively, N_{kz} and N_{ft} are the sampling numbers from the wavenumber domain to the space domain and that from the frequency domain to the time domain, respectively.

We point out two important observations in the RAI calculation. First, a smaller interval in the frequency domain gives more accurate result on the frequency domain integral, especially when the integrand contains rapid oscillation or singularities, as the acoustic logging response function in this work. Second, a smaller interval in the space domain guarantees a closer location to the point of interest for each waveform. The finite-difference grid size can affect the accuracy of the FDTD simulated waveforms. We found that the deviation becomes smaller than that in Fig. 3, when we reduce Δr and Δz by ten times, but this is costly.

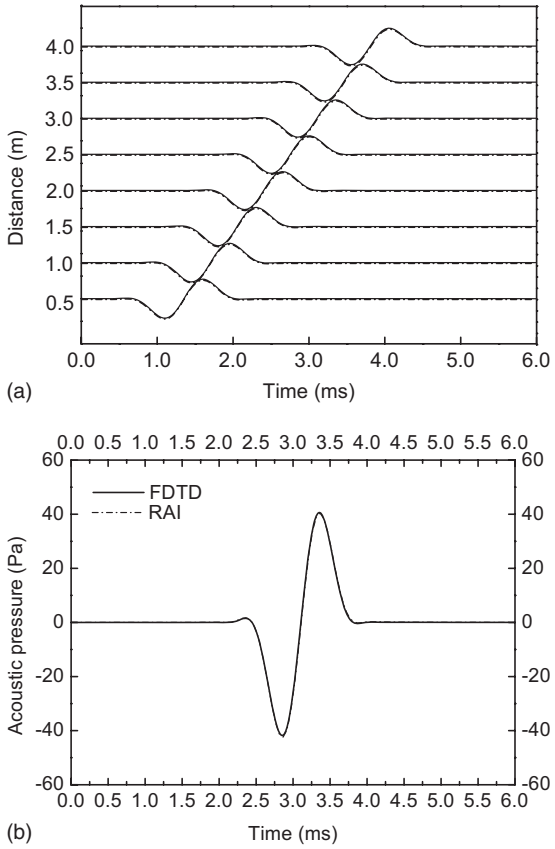


FIG. 4. Same as Fig. 3 except that the porous formation around the borehole is the Formation III and a center source frequency $f_0=1$ kHz is employed. (a) Pressure waveforms at different locations along the borehole axis. (b) The waveforms of (a) at the location $z=3.0$ m.

Figure 4 gives a comparison between the FDTD and the RAI methods of the simulated waveforms in a porous formation with a low permeability and a low porosity. The input parameters of the formation are listed in the second column (Formation II) of Table I. The formation has a compressional velocity of 4700 m/s and a shear velocity of 2990 m/s. The center frequency and the pulse width of the source are $f_0=1$ KHz and $T_c=2.0$ ms, respectively. At these frequencies, the compressional and shear waves are not very well excited, while the Stoneley wave dominates the waveforms. The finite-difference grid sizes are set to $\Delta r=\Delta z=0.01$ m and the time step is chosen as $\Delta t=1.5\times10^{-3}$ ms. Figure 4(a) shows excellent agreement between the two methods of the simulated waveforms at different locations along the borehole axis from $z=0.5$ m to $z=4.0$ m. Figure 4(b) shows more details at $z=3.0$ m.

In Figs. 3 and 4, the porous media surrounding the borehole are fast formations, where the shear velocity is faster than the borehole fluid velocity. The comparison for a slow porous formation is given in Fig. 5. The input parameters of the formation are listed in the third column (Formation III) of Table I. The formation has a compressional velocity of 2840 m/s and a shear velocity of 1190 m/s which is slower than the borehole fluid velocity. The critical frequency of this formation is $\omega_c=\phi\eta/\alpha_\infty\rho_f\kappa_0\approx1$ kHz. To check the validity of the FDTD algorithm for modeling poro-elastic wave problems with frequencies higher than the critical frequency, a

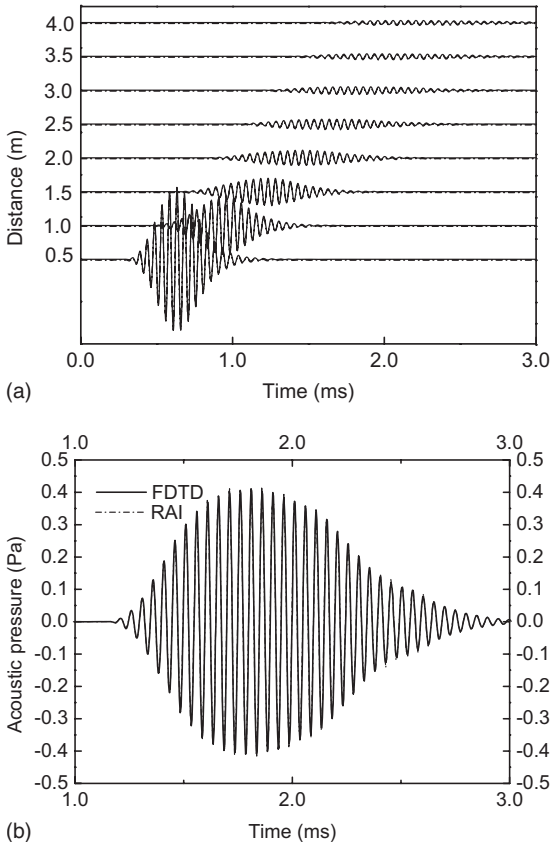


FIG. 5. Same as Fig. 3 except that the porous formation around the borehole is the Formation III and a higher source frequency $f_0=20$ kHz is employed. (a) Pressure waveforms at different locations along the borehole axis. (b) The waveforms of (a) at the location $z=3.0$ m.

source with a center frequency of $f_0=20$ kHz is employed. The finite-difference grid sizes are set to $\Delta r=\Delta z=0.0025$ m and the time step is chosen as $\Delta t=0.5\times10^{-3}$ ms. Figure 5(a) shows the comparison of the simulated pressure waveforms at different locations along the borehole axis. The compressional wave dominates the waveforms of the acoustic logging, because the shear and pseudo-Rayleigh wave cannot be received in soft formations and the Stoneley wave is not effectively excited at around 20 kHz. More details at $z=3$ m is given in Fig. 5(b). Note that the time axis of Fig. 5(b) is set from 1.0 to 3.0 ms, to show the phase of the waveforms clearly. Similar to the comparisons in Figs. 3 and 4, excellent agreement between the two methods is also observed in Fig. 5.

We have compared the FDTD modeled acoustic logs with the RAI method, with different source center frequencies from 1 to 20 kHz and in variety of formations with different velocities and permeabilities. The excellent agreement between the two methods of the waveforms in Figs. 3–5 confirms that the FDTD algorithm is valid and correct, and can be applied at frequencies higher than the critical frequency.

B. Horizontally stratified formation

Now we use the FDTD algorithm to simulate the acoustic logs in a horizontally stratified formation. The formation around the borehole has three layers, and the medium parameters of the layers are listed in Table I. The layer, which lies between 1.5 and 2.5 m above the source, is of Formation III (with high porosity and permeability). The other two layers are of Formation I.

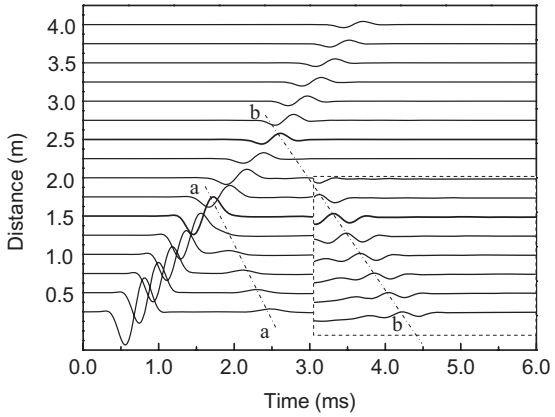


FIG. 6. FDTD modeled waveforms of the pressure at different locations along the borehole axis of the acoustic logs in a horizontally stratified porous formation. The layer, between 1.5 and 2.5 m, is the Formation III. The other two layers are the Formation I. The source center frequency is $f_0=2$ kHz.

Shown in Fig. 6 are the FDTD modeled pressure waveforms at different locations along the borehole axis. In order to show clearly the reflections of the Stoneley wave at the interfaces between two layers, the source center frequency $f_0=2$ kHz and the pulse width $T_c=2.0$ ms are employed. In Fig. 6, the line a-a marks the Stoneley wave reflected from the interface at $z=1.5$ m, which is similar to that in Kostek *et al.* (1998). In their work, the acoustic logs in elastic media with horizontal fluid-filled fractures was simulated. The reflected Stoneley wave from the second interface at $z=2.5$ m is very small. When we magnify by ten times the waveforms received from 3.0 to 6.0 ms, it is seen and is

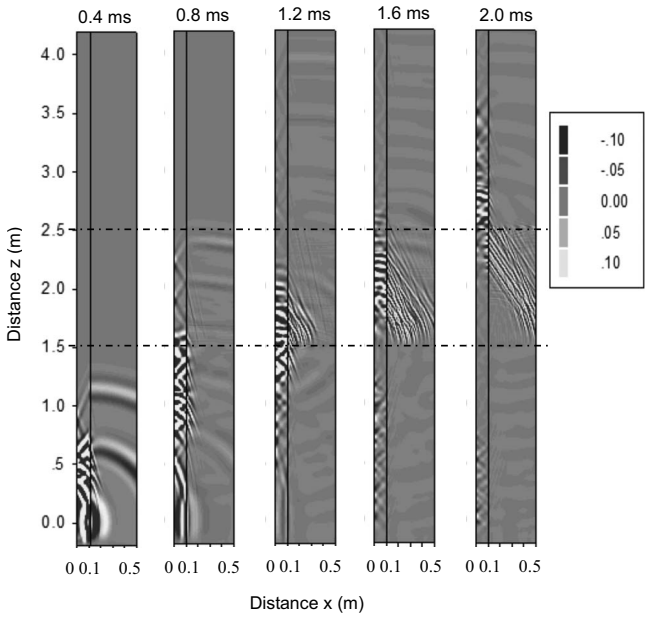


FIG. 7. Simulation of the wave-field distribution at five instants of 0.4, 0.8, 1.2, 1.6, and 2.0 ms of the acoustic logs in the horizontally stratified porous formation. The source center frequency is $f_0=20$ kHz.

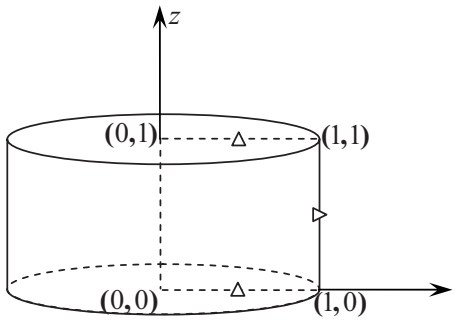


FIG. 8. Schematic of the volume integral region of the source-grid.

marked by the line *b*-*b* in Fig. 6. The reason is that most of the wave energy is attenuated in the interlayer and transmitted from the interface.

In Fig. 7, we show the wave-field distribution at instants of 0.4, 0.8, 1.2, 1.6, and 2.0 ms. In order to identify different wave groups in the wave-field, a higher center frequency $f_0=20$ kHz and a shorter pulse width $T_c=0.08$ ms are used in this simulation. In the slides of Fig. 7, no obvious wave reflection is observed from the outer boundaries, indicating that the wave-field is well absorbed in the PML. Shown in the first slide at 0.4 ms, the fastest pulse in the formation is the transmitted compressional wave from the borehole and then is the transmitted shear wave. The compressional head wave in the borehole also can be seen. In the slide at 0.8 ms, the large-amplitude arrival propagates to the interface at $z=1.5$ m, which is the pseudo-Rayleigh wave in the borehole. For the frequency range in this simulation, the Stoneley wave is too weak to be identified within the pseudo-Rayleigh wave packet. In the three later slides at 1.2, 1.6, and 2.0 ms, it is seen that part of the pseudo-Rayleigh wave energy in the borehole is reflected back at the interface, and part of the wave energy is leaked into the porous formation by the Biot slow wave (Biot 1956a, 1956b) and is rapidly dissipated. The Biot slow wave is related to the fluid motion in the porous formation, whose velocity is around 650 m/s in the interlayer. It moves much deeper into the formation at the location from $z=1.5$ m to $z=2.5$ m, because the attenuation coefficient in the interlayer is eight times lower than that in the other two layers.

IV. CONCLUSIONS

We have proposed a velocity-stress FDTD algorithm with nonsplitting PML to model the monopole acoustic logs in horizontally stratified fluid-saturated porous formations. By considering the inertial force on the viscous fluid flow in pore space, the algorithm can be applied in cases with frequencies higher than the critical frequency. In acoustic logging, the discretization of the field quantities on the borehole wall is difficult than that in seismic exploration, because the governing equations in the borehole are different from those in the formation. We have discretized the field quantities on the borehole wall in Appendix B.

The proposed FDTD algorithm is used to simulate axisymmetric acoustic logging responses excited by sources with center frequencies from 1 to 20 kHz and in homogeneous formations with different velocities and permeabilities. The FDTD simulated waveforms are compared with those by the RAI methods. The excellent agreement in every comparison confirms the validity and correctness of the FDTD algorithm.

In an example for the horizontally stratified formation, the reflected Stoneley waves from the upper and lower interfaces of the interlayer are observed in the waveforms, while the transmission wave becomes small in amplitude. An interesting phenomenon is the appearance of Biot slow wave in the soft interlayer, which originates on the borehole wall and propagates into the formation.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (Grant No. 40874062) and the Special Research Funds of Seismology in China (Grant No. 200808072). The authors are very grateful to Yubao Zhen, Ralph Stephen, and the anonymous reviewers for their revision suggestions for this paper.

APPENDIX A: TREATMENT OF THE POINT SOURCE AND DISCRETIZATION OF THE EQUATIONS IN THE BOREHOLE

The velocity-stress equations of the acoustic wave in the fluid-filled borehole can be expressed as (Liu and Tao, 1997)

$$\rho_b \frac{\partial \mathbf{v}_f}{\partial t} = -\nabla p_f, \tag{A1}$$

$$\frac{\partial p_f}{\partial t} = -\rho_b c_{bm}^2 \nabla \cdot \mathbf{v}_f + F(r,z,t), \tag{A2}$$

where p_f is the borehole fluid pressure, $\mathbf{v}_f$ and ρ_b are the velocity vector and the density of the borehole fluid, respectively, c_{bm} is the sound speed in the borehole fluid, and $F(r,z,t)$ is the function determined by the source, which becomes zero for source-free grids.

In terms of the staggered grid shown in Fig. 2 and noting that Ω_r and $\bar{\Omega}_r$ become zero due to the absence of the r-direction PML in the borehole, the acoustic wave equations for source-free grids are discretized as

$$v_{fr(j,k+1/2)}^{n+1/2} = v_{fr(j,k+1/2)}^{n-1/2} - \frac{\Delta t}{\rho_b} \frac{\partial p_{f(j,k+1/2)}^n}{\partial r}, \tag{A3}$$

$$v_{fz(j+1/2,k)}^{n+1/2} = v_{fz(j+1/2,k)}^{n-1/2} - \frac{\Delta t}{\rho_b} \left(\frac{\partial p_f^n}{\partial z} + P_{pfz}^n \right)_{(j+1/2,k)}, \tag{A4}$$

$$p_{f(j+1/2,k+1/2)}^{n+1} = p_{f(j+1/2,k+1/2)}^n - \rho_b c_{bm}^2 \Delta t \left[\frac{\partial v_{fr}^{n+1/2}}{\partial r} + \frac{v_{fr}^{n+1/2}}{r} + \left(\frac{\partial v_{fz}^{n+1/2}}{\partial z} + P_{vfz}^{n+1/2} \right) \right]_{(j+1/2,k+1/2)}, \tag{A5}$$

where v_{fr} and v_{fz} are the components of the velocity vector $\mathbf{v}_f$, and

$$P_{pfz}^n = e^{-\Omega_z \Delta t} P_{vfz}^{n-1} - \frac{1}{2} \Omega_z \Delta t \left(e^{-\Omega_z \Delta t} \frac{\partial p_f^{n-1}}{\partial z} + \frac{\partial p_f^n}{\partial z} \right), \tag{A6}$$

$$P_{vfz}^{n+1/2} = e^{-\Omega_z \Delta t} P_{vfz}^{n-1/2} - \frac{1}{2}\Omega_z \Delta t \left(e^{-\Omega_z \Delta t} \frac{\partial v_{fz}^{n-1/2}}{\partial z} + \frac{\partial v_{fz}^{n+1/2}}{\partial z} \right). \tag{A7}$$

The point source employed is the same as described by Tsang and Rader (1979), with the peak pressure being 100 Pa at a location of 0.01 m away from the source. This source is introduced by letting $F(r,z,t)=4\pi c_{bm}^2 \delta(r,z)S(t)$ in Eq. (A2), where $\delta(r,z)$ is the delta function, $S(t)=\int_0^t s(t')dt'$, and the source pulse function $s(t)$ used in this paper is

$$s(t) = \begin{cases} \frac{1}{2}\left[1+\cos\frac{2\pi}{T_c}\left(t-\frac{T_c}{2}\right)\right]\cos 2\pi f_0\left(t-\frac{T_c}{2}\right), & 0 \le t \le T_c \\ 0, & t<0 \text{ or } t>T_c, \end{cases} \tag{A8}$$

where f_0 and T_c denote the center frequency and the pulse width of the source, respectively. Equation (A2) for the source-grid can be discretized by the volume integral method. Using the generalized Gauss formula, the volume integral of Eq. (A2) is expressed as (see Fig. 8)

$$\int\int\int_\Omega \frac{\partial p_f}{\partial t} dV = -\rho_b c_{bm}^2 \oiint_\Sigma \mathbf{v}_f \cdot d\mathbf{s} + 4\pi c_{bm}^2 S(t) \int\int\int_\Omega \delta(r,z) dV, \tag{A9}$$

where Ω and Σ denote the volume and the surface area of the cylindrical body shown in Fig. 8. Applying the integral mean value theorem to Eq. (A9) yields

$$p_{f(1/2,1/2)}^{n+1} = p_{f(1/2,1/2)}^{n} + \frac{c_{bm}^2 \Delta t}{\Delta z}\left[-\rho_b \left(v_{fz(1/2,1)}^{n+1/2} - v_{fz(1/2,0)}^{n+1/2} + 2\frac{\Delta z}{\Delta r} v_{fr(1,1/2)}^{n+1/2} \right) + \frac{4}{(\Delta r)^2} F^{n+1/2} \right]. \tag{A10}$$

APPENDIX B: TREATMENT OF THE FIELD QUANTITIES AT THE INTERFACE

On the borehole wall ($j=r_a$), the quantity $v_{fr(r_a,k+1/2)}^{n+1/2}$ cannot be discretized directly according to Eq. (A3), because the space derivative $\partial/\partial r$ in Eq. (A3) breaks at the medium interface. Using the approximation $\partial p_{f(r_a,k+1/2)}^n/\partial r \approx \partial p_{f(r_a-1/4,k+1/2)}^n/\partial r$ in Eq. (A3) yields

$$v_{fr(r_a,k+1/2)}^{n+1/2} = v_{fr(r_a,k+1/2)}^{n-1/2} - \frac{\Delta t}{\rho_b} \frac{\partial p_{f(r_a-1/4,k+1/2)}^n}{\partial r}. \tag{B1}$$

Similarly, the quantities $v_{ur(r_a,k+1/2)}^{n+1/2}$ and $v_{wr(r_a,k+1/2)}^{n+1/2}$ on the borehole wall cannot be obtained from Eqs. (12) and (15). Introducing the approximation $\partial p_{(r_a,k+1/2)}^n/\partial r \approx \partial p_{(r_a+1/4,k+1/2)}^n/\partial r$ to Eq. (12) and the approximations $\partial \tau_{rr(r_a,k+1/2)}^n/\partial r \approx \partial \tau_{rr(r_a+1/4,k+1/2)}^n/\partial r$ and $\tau_{\theta\theta(r_a,k+1/2)}^n \approx (3\tau_{\theta\theta(r_a+1/2,k+1/2)}^n - \tau_{\theta\theta(r_a+3/2,k+1/2)}^n)/2$ to Eq. (15), and using the fact that the variables P_{pr}^n in Eq. (12) and P_{rrr}^n and $Q_{r\_\theta r}^n$ in Eq. (15) are zero on the borehole wall, we obtain

$$\frac{\eta}{\kappa_0}\left[\frac{1}{2}+\left(\frac{D_1}{2}+D_2\right)\frac{1}{\Delta t}\right] v_{wr(r_a,k+1/2)}^{n+1/2} + \frac{\rho_f}{\Delta t} v_{ur(r_a,k+1/2)}^{n+1/2} = \frac{\eta}{\kappa_0}\left[-\frac{1}{2}+\left(\frac{D_1}{2}+D_2\right)\frac{1}{\Delta t}\right] v_{wr(r_a,k+1/2)}^{n-1/2} + \frac{\rho_f}{\Delta t} v_{ur(r_a,k+1/2)}^{n-1/2} - \frac{\partial p_{(r_a+1/4,k+1/2)}^n}{\partial r}, \tag{B2}$$

$$\rho v_{ur(r_a,k+1/2)}^{n+1/2} + \rho_f v_{wr(r_a,k+1/2)}^{n+1/2} = \rho v_{ur(r_a,k+1/2)}^{n-1/2} + \rho_f v_{wr(r_a,k+1/2)}^{n-1/2} + \Delta t \left[\frac{\partial \tau_{rr(r_a+1/4,k+1/2)}^n}{\partial r} - \frac{1}{r}\left(\frac{3}{2}\tau_{\theta\theta(r_a+1/2,k+1/2)}^n - \frac{1}{2}\tau_{\theta\theta(r_a+3/2,k+1/2)}^n \right) + \left(\frac{\tau_{rr}^n}{r} + \frac{\partial \tau_{rz}^n}{\partial z} + P_{rzz}^n \right)_{(r_a,k+1/2)} \right]. \tag{B3}$$

The acoustic fields on the borehole wall obey the boundary conditions of the open pore interface (Hu *et al.* 2007):

$$v_{fr(r_a,k+1/2)}^{n+1/2} = v_{ur(r_a,k+1/2)}^{n+1/2} + v_{wr(r_a,k+1/2)}^{n+1/2}, \tag{B4}$$

$$p_{f(r_a,k+1/2)}^n = p_{(r_a,k+1/2)}^n, \tag{B5}$$

$$-p_{f(r_a,k+1/2)}^n = \tau_{rr(r_a,k+1/2)}^n, \tag{B6}$$

$$0 = \tau_{rz(r_a,k)}^n. \tag{B7}$$

From Eq. (B7), we derive that the term $(\partial \tau_{rz}^n/\partial z + P_{rzz}^n)_{(r_a,k+1/2)}$ in Eq. (B3) is zero. When substituting Eqs. (B5) and (B6) into Eqs. (B2) and (B3), respectively, there remain four unknown field quantities in Eqs. (B1)–(B4). By solving these equations, we can derive the field quantities on the borehole wall, i.e., $v_{fr(r_a,k+1/2)}^{n+1/2}$, $v_{ur(r_a,k+1/2)}^{n+1/2}$, and $v_{wr(r_a,k+1/2)}^{n+1/2}$.

Alford, R. M., Kelly, K. R., and Boore, D. M. (**1974**). "Accuracy of finite-difference modeling of the acoustic wave equation," Geophysics **39**, 834–842.

Berenger, J. P. (**1994**). "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200.

Biot, M. A. (**1956a**). "Theory of propagation of elastic waves in a fluid-

saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178.
Biot, M. A. (**1956b**). "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191.
Biot, M. A. (**1962**). "Mechanics of deformation and acoustic propagation in porous media," J. Appl. Phys. **33**, 1482–1498.
Cerjan, C., Kosloff, D., Kosloff, R., and Reshef, M. (**1985**). "A nonreflecting boundary condition for discrete acoustic and elastic wave equation," Geophysics **50**, 705–708.
Chew, W. C., and Liu, Q. H. (**1996**). "Perfectly matched layers for elastodynamics: A new absorbing boundary condition," J. Comput. Acoust. **4**, 341–359.
Chew, W. C., and Weedon, W. H. (**1994**). "A 3D perfectly matched medium from modified Maxwell's equations with stretched coordinates," Microwave Opt. Technol. Lett. **7**, 599–604.
Clayton, R., and Engquist, B. (**1977**). "Absorbing boundary conditions for acoustic and elastic wave equations," Bull. Seismol. Soc. Am. **67**, 1529–1540.
Dai, N., Vafidis, A., and Kanasewich, E. R. (**1995**). "Wave propagation in heterogeneous, porous media: A velocity-stress, finite-difference method," Geophysics **60**, 327–340.
Dong, H., Kaynia, A. M., Madshus, C., and Hovern, J. M. (**2000**). "Sound propagation over layered poro-elastic ground using a finite-difference model," J. Acoust. Soc. Am. **108**, 494–502.
Guan, W., and Hu, H. (**2008**). "Finite-difference modeling of the electroseismic logging in a fluid-saturated porous formation," J. Comput. Phys. **227**, 5633–5648.
Hardin, E. L., Cheng, C. H., Paillet, F. L., and Mendelson, J. D. (**1987**). "Fracture characterization by means of attenuation and generation of tube waves in fractured crystalline rock at Mirror Lake, New Hampshire," J. Geophys. Res. **92**, 7989–8006.
Hastings, F. D., Schneider, J. B., and Broschat, S. L. (**1996**). "Application of the perfectly matched layer PML absorbing boundary condition to elastic wave propagation," J. Acoust. Soc. Am. **100**, 3061–3069.
Higdon, R. L. (**1991**). "Absorbing boundary conditions for elastic waves," Geophysics **56**, 231–241.
Hornby, B. E. (**1989**). "Imaging near-borehole of formation structure using full-waveform sonic data," Geophysics **54**, 747–757.
Hornby, B. E., Johnson, D. L., Winkler, K. W., and Plumb, R. A. (**1989**). "Fracture evaluation using reflected Stoneley-wave arrivals," Geophysics **54**, 1274–1288.
Hu, H., Guan, W., and Harris, J. M. (**2007**). "Theoretical simulation of electroacoustic borehole logging in a fluid-saturated porous formation," J. Acoust. Soc. Am. **122**, 135–145.
Johnson, D. L., Koplik, J., and Dashen, R. (**1987**). "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402.
Kosloff, R., and Kosloff, D. (**1986**). "Absorbing boundary for wave propagation problems," J. Comput. Acoust. **63**, 363–376.
Kostek, S., Johnson, D. L., and Randall, C. J. (**1998**). "The interaction of tube waves with borehole fractures, Part I: Numerical models," Geophysics **63**, 800–808.
Liao, Z. P., Wong, H. L., Yang, B. P., and Yuan, Y. F. (**1984**). "A transmitting boundary for transient wave analysis," Sci. Sin., Ser. A **27**, 1063–1076.
Liu, Q. H., and Tao, J. P. (**1997**). "The perfectly matched layer for acoustic waves in absorptive media," J. Acoust. Soc. Am. **102**, 2072–2082.
Masson, Y. J., Pride, S. R., and Nihei, K. T. (**2006**). "Finite difference modeling of Biot's poroelastic equations at seismic frequencies," J. Geophys. Res. **111**, B10305.
Peng, C. B., and Toksöz, M. N. (**1995**). "An optimal absorbing boundary condition for elastic wave modeling," Geophysics **60**, 296–301.
Reynolds, A. C. (**1978**). "Boundary conditions for the numerical solution of wave propagation problems," Geophysics **43**, 1099–1110.
Roden, J. A., and Gedney, S. D. (**2000**). "Convolution PML(CPML): An efficient FDTD implementation of the CFS-PML for arbitrary media," Microwave Opt. Technol. Lett. **27**, 334–339.
Rosenbaum, J. H. (**1974**). "Synthetic microseismograms: Logging in porous formations," Geophysics **39**, 14–32.
Schmitt, D. P. (**1988**). "Effects of radial layering when logging in saturated porous formations," J. Acoust. Soc. Am. **84**, 2200–2214.
Schmitt, D. P., Bouchon, M., and Bonnet, G. (**1988**). "Full-wave synthetic acoustic logs in radially semiinfinite saturated porous media," Geophysics **53**, 807–823.
Sochacki, J., Kubichek, R., George, J., Fletcher, W. R., and Smithson, S. (**1987**). "Absorbing boundary conditions and surface waves," Geophysics **52**, 60–71.
Taflove, A., and Hagness, S. C., (**2005**). *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed. (Artech House, Norwood, MA), p. 75.
Tang, X. M., and Cheng, C. H. (**1996**). "Fast inversion of formation permeability from Stoneley wave logs using a simplified Biot–Rosenbaum model," Geophysics **61**, 639–645.
Tang, X. M., Zheng, Y., and Patterson, D. (**2007**). "Processing array acoustic-logging data to image near-borehole geologic structures," Geophysics **72**, E87–E97.
Tsang, L., and Rader, D. (**1979**). "Numerical evaluation of the transient acoustic waveform due to a point source in a fluid-filled borehole," Geophysics **44**, 1706–1720.
Vernik, L. (**1994**). "Predicting lithology and transport properties from acoustic velocities based on petrophysical classification of siliclastics," Geophysics **59**, 420–427.
Virieux, J. (**1986**). "*P-SV* wave propagation in heterogeneous media: Velocity-stress finite-difference method," Geophysics **51**, 889–1001.
Wang, T., and Tang, X. M. (**2003**). "Finite-difference modeling of elastic wave propagation: A nonsplitting perfectly matched layer approach," Geophysics **68**, 1749–1755.
Williams, D. M., Zemanek, J., Angona, F. A., Dennis, C. L., and Caldwell, R. L. (**1984**). "The long space acoustic logging tool," *25th Annual Logging Symposium Transactions*, Society of Professional Well Log Analysts, New Orleans, LA, Paper T.
Zeng, Y. Q., He, J. Q., and Liu, Q. H. (**2001**). "The application of the PML in numerical modeling of wave propagation in poroelastic media," Geophysics **66**, 1258–1266.
Zeng, Y. Q., and Liu, Q. H. (**2001**). "A staggered-grid finite-difference method with perfectly matched layers for poroelastic wave equations," J. Acoust. Soc. Am. **109**, 2571–2580.
Zhu, X., and McMechan, G. A. (**1991**). "Numerical simulation of seismic responses of poroelastic reservoirs using Biot theory," Geophysics **56**, 328–339.

Laboratory investigation of the acoustic response of seagrass tissue in the frequency band 0.5–2.5 kHz

Preston S. Wilson[a)]
Department of Mechanical Engineering and Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 77712-0292

Kenneth H. Dunton
Marine Science Institute, The University of Texas at Austin, Port Aransas, Texas 78373-5015

(Received 2 October 2007; revised 29 January 2009; accepted 30 January 2009)

Previous *in situ* investigations of seagrass have revealed acoustic phenomena that depend on plant density, tissue gas content, and free bubbles produced by photosynthetic activity, but corresponding predictive models that could be used to optimize acoustic remote sensing, shallow water sonar, and mine hunting applications have not appeared. To begin to address this deficiency, low frequency (0.5–2.5 kHz) acoustic laboratory experiments were conducted on three freshly collected Texas Gulf Coast seagrass species. A one-dimensional acoustic resonator technique was used to assess the biomass and effective acoustic properties of the leaves and rhizomes of *Thalassia testudinum* (turtle grass), *Syringodium filiforme* (manatee grass), and *Halodule wrightii* (shoal grass). Independent biomass and gas content estimates were obtained via microscopic cross-section imagery. The acoustic results were compared to model predictions based on Wood's equation for a two-phase medium. The effective sound speed in the plant-filled resonator was strongly dependent on plant biomass, but the Wood's equation model (based on tissue gas content alone) could not predict the effective sound speed for the low irradiance conditions of the experiment, in which no free bubbles were generated by photosynthesis. The results corroborate previously published results obtained *in situ* for another seagrass species, *Posidonia oceanica*. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3086272]

PACS number(s): 43.30.Ma [KGF] Pages: 1951–1959

I. INTRODUCTION

The acoustical characteristics of seagrass beds are important in a number of naval, commercial, and ecological applications. Gas-filled channels (aerenchyma) within the seagrass plants and gas bubbles generated by the plants during photosynthesis dominate their acoustic behavior and numerous studies have demonstrated and analyzed significant acoustic effects due to this acoustic contrast with the water and ocean bottom. For example, mine hunting performance is severely reduced in seagrass meadows. McCarthy and Sabol[1] attempted to locate a Manta target simulator (an inert version of a Manta naval mine) that was deployed in a meadow of *Zostera marina* using two commercial side scan sonars operating at both 100 and 500 kHz. Neither sonar was able to detect the target in numerous attempts. Similarly, a hand-held mine hunting sonar deployed by divers and operating in the 115–140 kHz frequency range was unable to detect the target.[1] Backscatter from the seagrass was sufficiently strong to completely obscure reflections from the Manta target simulator.

Since shallow water sonar performance can be severely limited by reverberation, Lyons and Abraham[2] studied the variation in seafloor reverberation level at 80 kHz in the presence of *Posidonia oceanica* near Sardinia and Sicily. They found that the mean seafloor backscatter strength in the presence of seagrass was about 10 dB higher across a range of grazing angles than compared to a sand-covered bottom and varied between 0 and 8 dB higher than a mud bottom, depending on grazing angle. While the mean backscatter strength was not uniquely related to the bottom characteristics, they found that a combination of mean backscatter strength and a statistical description of the variability of the backscatter could potentially characterize the bottom. The variability was attributed to the inhomogeneous distribution of *Posidonia* density and the motion of leaves due to swell or currents.

Seagrass meadows serve as habitat for many economically important estuarine organisms. The same acoustic characteristics that obscure mines and hinder sonar performance can potentially be used to assess the ecological condition and the extent of this valuable littoral ecosystem. Consequently, acoustic remote sensing methods are under development to exploit these characteristics.[3] Side scan sonar has been used extensively for imaging seagrass beds and as a possible management tool for littoral ecosystems.[4] An extensive comparison between seagrass mapping by side scan sonar and by aerial photography was conducted by Mulhearn.[5] Komatsu *et al.*[6] showed that plant-occupied volume can be extracted from the three-dimensional images produced by multi-beam side scan sonar. Finally, Hermand[3] conducted extensive studies over the past decade on the acoustic remote sensing of seagrass photosynthesis using frequencies ranging from 0.1 to 16 kHz. One study[7] on *Posidonia oceanica* was conducted

[a)]Electronic mail: pswilson@mail.utexas.edu

in the frequency range 0.1–1.6 kHz, which overlaps the frequency range used in this work. Photosynthetic activity produces free gas bubbles on the plants and in the water column which affects the local sound speed and hence the waveguide impulse response. Hermand's technique relies on monitoring the acoustic impulse response between two locations in a shallow water waveguide and inverting for a measurement of oxygen productivity.

In all the applications described above, accurate predictive models that relate the physical characteristics of plant anatomy, density, and photosynthetic activity to the acoustic scattering and propagation response could potentially be used to (1) infer target and ecological parameters from acoustic sonar and remote sensing signals and (2) improve and optimize acoustic image formation for mapping and target identification. As a first step toward the development of such models, the hypothesis that the acoustic behavior of immersed plant leaves or rhizomes can be considered an effective medium dominated by internal plant gas content was investigated. Photosynthetically produced gas bubbles attached to tissue surfaces or in the water surrounding the plants were excluded. Low frequency (0.5–2.5 kHz) laboratory experiments were conducted on three freshly collected species: *Thalassia testudinum* (turtle grass), *Syringodium filiforme* (manatee grass), and *Halodule wrightii* (shoal grass). This frequency range is directly applicable to the remote sensing application of Hermand[3] and Hermand *et al.*[7] but is significantly lower than the frequencies used in side scan sonar studies and object detection applications. Despite this discrepancy, the frequency range used in the present study is an appropriate starting point for the investigation of the fundamental physics of a fluid-loaded gas-bearing structure such as seagrass tissue because at sufficiently low frequencies, below the resonance frequency of any individual gas-filled cavity, the relative acoustic importance of the gas content compared to the tissue can be assessed through the use of an effective medium model, such as Wood's equation.[8] This study is not intended to lead directly to a model of propagation and scattering in the presence of seagrass in the frequency range of side scan and mine hunting sonar systems, but is instead intended to yield basic insight into the acoustic behavior of seagrass tissue.

A one-dimensional acoustic resonator technique was used to assess the biomass and effective acoustic properties of individual leaves and rhizomes. Independent biomass and gas content estimates were gathered using macro- and microscopic tissue imagery. The acoustic results were compared to model predictions based on Wood's equation[8] for a two-phase medium. The effective sound speed in the plant-filled resonator was strongly dependent on plant biomass, but the Wood's equation model (based on gas content alone) was unable to describe the effective sound speed for the low irradiance conditions of the experiment, in which no free bubbles were generated by photosynthesis. These laboratory results corroborated findings obtained *in situ* for another seagrass species, *Posidonia oceanica*.[3]

II. LOW FREQUENCY ACOUSTIC RESPONSE AND PLANT BIOMASS

Seagrass leaves and rhizomes contain internal gas-filled channels (aerenchyma)[9] and during photosynthesis, gas bubbles produced by the plant are found attached to external surfaces. This study was focused on the effects of the aerenchyma on the acoustic response characteristics; hence the experiments were done in sufficiently low light conditions to prevent external bubble formation. The following hypotheses were investigated: (1) the acoustic response is dominated by the gas within the aerenchyma and (2) seagrass tissues, when gas-free, possess acoustic properties close to those of water. Under these assumptions, the contents of the acoustic resonator (seawater and plants) were considered to be an effective acoustic medium composed of a liquid seawater phase and a gas phase, as is done for air bubbles in water.

The speed of sound in this medium is a strong function of the gas content, or void fraction, which is defined as $\chi = V_{\mathrm{gas}}/V_{\mathrm{tot}}$. The volume of gas in the aerenchyma is V_{gas} and the total volume of the resonator is V_{tot}. If the frequency of the acoustic excitation is well below the resonance frequency of the largest gas cavity, then the effective medium sound speed c_{eff} is given by Wood's equation[8]

$$\frac{1}{c_{\mathrm{eff}}^2} = \frac{(1-\chi)^2}{c_\ell^2} + \frac{\chi^2}{c_g^2} + \chi(1-\chi)\frac{\rho_g^2 c_g^2 + \rho_\ell^2 c_\ell^2}{\rho_\ell \rho_g c_\ell^2 c_g^2}, \tag{1}$$

where the subscripts ℓ and g refer to the liquid and gas phases, respectively, c_i is the intrinsic sound speed of phase i, and ρ_i is the density of phase i. If hypotheses (1) and (2) are true, then the sound speed observed inside the resonator should be directly related to the aerenchyma gas volume via Eq. (1) and also related to the plant biomass. In this study, leaf and rhizome volume was used to represent the biomass. The exact composition of the gas inside the aerenchyma is not known. Since it must contain oxygen and potentially nitrogen, the gas was considered to be air. The density difference between oxygen and nitrogen leads to a sound speed difference between the pure gases of about 7%, but this was found to be negligible in comparison to other sound speed effects observed in this study. The sound speed for air (315 m/s) originally cited by Wood was used,[8] which is appropriate for air bubbles in water at the surface.[10]

III. EXPERIMENT

Acoustical response trials were conducted *in vitro* using a one-dimensional acoustic resonator (Fig. 1) constructed from transparent polyvinyl chloride (PVC) schedule 40 pipe of circular cross-section (60 ± 0.5 mm outer diameter and 4.2 ± 0.2 mm wall thickness) and with length $L = 401 \pm 1$ mm. The cylinder was filled with fresh seawater from the Gulf of Mexico that had spent several hours in an outdoor settling tank to minimize interference from suspended particulates. Seagrass leaves or rhizomes were placed in the resonator (additional details about the plant tissue preparation are provided below). The air-water interface at the top and a Styrofoam block at the bottom of the cylinder provided pressure release acoustic boundary conditions to a high degree of approximation, as described in Ref. 11.

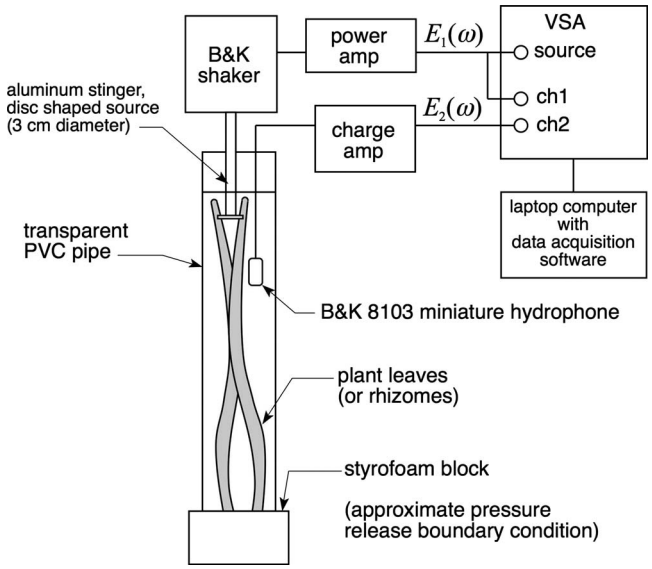


FIG. 1. The acoustic resonator and the measurement instrumentation are shown.

Acoustic standing waves were generated with a 3-cm-diameter aluminum piston that was attached to an electromechanical shaker via an aluminum stinger. Band-limited periodic chirps were produced by a vector signal analyzer (VSA) and directed to the shaker through a power amplifier. The source signal was also digitized by the VSA. The piston was positioned a few centimeters below the air-water interface. The acoustic pressure inside the resonator was received with a miniature hydrophone, positioned a few centimeters below the source. The received signals were amplified with a charge amplifier, band-pass filtered (10 Hz–10 kHz), and digitized by the VSA. According to manufacturer specifications, the source velocity response was flat to within less than ± 1 dB in the experimental frequency range.

Relative acoustic pressure spectra y were then calculated onboard the VSA by way of a transfer function $y = P(\omega)/E_1(\omega) = [E_2(\omega)/M]/E_1(\omega)$, which yields the acoustic pressure $P(\omega)$ per unit voltage input $E_1(\omega)$ (see Fig. 1). The hydrophone system sensitivity M was used to convert the voltage $E_2(\omega)$ measured at the output of the charge amplifier into an acoustic pressure $P(\omega)$. According to the manufacturer's calibration, M varied less than ± 1 dB within the experimental frequency range and it was taken to be a constant. The spectra were calculated using 1601 frequency bins, 1.5 Hz resolution bandwidth, and 20 spectral averages, as defined by Eqs. (24)–(27) of Ref. 12. The use of a transfer function measurement preserved the phase of the acoustic pressure relative to the excitation signal and yielded the coherence function. The number of averages and the amplitude of the excitation were chosen to achieve a near-unity coherence function at all frequencies of interest, which guaranteed a high signal-to-noise ratio and ensured linear behavior.[13] Peak pressures inside the resonator were typically less than 170 dB re 1 μPa.

Live plants were collected intact with sediment cores from the Gulf of Mexico near the Marine Science Institute (MSI) in Port Aransas, TX and transported in seawater to MSI. *Thalassia* plants were collected on 9 November 2005, and *Halodule* and *Syringodium* plants were collected on 14 August 2007. The living plant cores were stored in transparent cylindrical chambers filled with seawater for 12–24 h between collection and the acoustic measurements. Previous work has shown that this artificial environment causes no detrimental effects on the plants within 5 days of collection.[14] The experimental protocol described below refers specifically to the preparation of leaves, but the same procedures were followed for rhizomes. The desired number of leaves was first removed from a plant by hand, rinsed of sediment in a seawater bath, and then transferred to the acoustic resonator, all without exposing the leaves to the air. No direct lighting was used. Only the diffuse fluorescent lighting present in the laboratory was incident upon the plants and no photosynthetic gas production expressed through external bubble formation was observed during the experiments. Hence these experiments represent a baseline case, corresponding to a minimum level of photosynthetic activity. It is known that the plant gas content and external bubble formation are a function of photosynthetic activity,[3] but variation in the illumination level was not included in this study.

The acoustic pressure spectra were then obtained as described above. The temperature of the water within the resonator was measured with a thermocouple gauge and ranged from 20.5 to 25.6 °C, depending on the time of day at which the seawater was drawn from the outdoor settling tank. The salinity was measured with a refractometer and remained constant throughout the experiments at 30 ppt. Temperature and salinity measurements were used to calculate the density and sound speed of the water using the UNESCO International Equation of State.[15] The mean water temperature near Port Aransas, TX, where these plants were collected, ranges from 13 to 30 °C over the course of the year;[16] hence the laboratory temperatures are within the range encountered in nature.

The leaves were removed from the resonator, laid flat on a table, and photographed alongside a machinist's scale. The surface area S of each leaf was measured using image analysis software. The thickness t was measured by imaging approximately five excised cross-sections from each leaf with a video microscope. The volume V_{leaf} of each leaf was then obtained from $V_{\text{leaf}} = S \times \bar{t}$, where $\bar{t}$ is the mean thickness of each leaf. Since the rhizomes of all three species and the leaves of *Syringodium* are circular in cross-section, their volumes were calculated from length (l) and the projected area (S_p) measurements using image analysis software; rhizome or leaf volume was calculated from $V = \pi S_p^2/4l$.

The leaf void fraction was obtained from image analysis that was performed on microscopic leaf cross-sections from the literature,[17] but no such cross-section images were found in the literature for the rhizomes. Digital cross sectional images of rhizomes were obtained using a microscope with backlighting. In both cases, image analysis was used to obtain the total area of the cross-section A_{tot} and the combined area of all the pores A_p within the cross-section (which are distinct in the images). Since the cross-sections are nearly uniform in the direction normal to the plane of the cross-

TABLE I. A different resonator tube was used for each seagrass species. The physical properties used in the evaluation of the elastic waveguide model for each resonator are shown. For all three, the resonator tube material was schedule 40 PVC pipe. When two entries appear within a cell, they are for the rhizomes and leaves, respectively.

	Thallasia	*Syringodium*	*Halodule*
Seawater sound speed, c_0 (m/s)	1517	1521	1517, 1519
Seawater density, ρ_1 (kg/m^3)	1020.9	1019.4	1017.3
Seawater temperature (°C)	20.5	25.6	22, 23
Air sound speed, c_g (m/s)	315	315	315
Air density, ρ_g (kg/m^3)	1.21	1.21	1.21
PVC comp. sound speed, c_l (m/s)	2030	2020	2020
PVC Poisson's ratio, ν	0.38	0.38	0.38
PVC density, ρ_w (kg/m^3)	1330	1330	1330
Inner wall radius, b (cm)	2.58	2.60	2.60
Outer wall radius, d (cm)	3.06	3.01	3.01

section images, the leaf void fractions were obtained with $\chi_{\text{leaf}}=V_{\text{air}}/V_{\text{leaf}}=A_p l/A_{\text{tot}} l=A_p/A_{\text{tot}}$. The same was done for the rhizomes. The existence of the leaf cross-sections in the literature was fortunate because sectioning the leaf tissue without causing visible damage to the leaf structure was difficult. The structures inside the leaves are generally smaller than the structures inside the rhizomes, and the leaf tissue is also softer than the corresponding rhizome tissue and tears more easily. Variation in leaf and rhizome internal void fraction within the species is possible, but to the author's knowledge, this has not been quantified in the literature.

IV. ELASTIC WAVEGUIDE EFFECT

In a gas-filled acoustic resonator, the walls are effectively rigid. In a liquid-filled resonator, there is significant coupling between the fill-liquid and the resonator walls. The result is a reduced sound speed relative to that observed in an unconfined environment. This effect is hereafter referred to as the elastic waveguide effect. An exact analytic model[18] for sound propagation in a finite-thickness elastic-walled, fluid-filled cylindrical tube was used to relate the speeds observed in the resonator to the intrinsic sound speed the material would exhibit in an unconfined environment. This procedure (and its validity) is discussed in Ref. 11 and summarized here. Equation (A1) of Ref. 11 is the dispersion relation for the resonator waveguide. The intrinsic (free-field) sound speed c_0 of the material that fills the resonator is an input parameter to Eq. (A1). The phase speed c_{ph} of the resonator's plane wave mode is an output. The resonator measurements described here yield the effective phase speed c_{eff} observed inside the resonator. The intrinsic sound speed c_0 is then varied in Eq. (A1) until the model output c_{ph} matches the measured value c_{eff}. The value of c_0 that achieves the match is reported as the effective sound speed that the material inside the resonator (seawater and plants) would exhibit in the ocean.

Three replicate resonator tubes were used in this work. The physical parameters needed for the elastic waveguide model [Eq. (A1), Ref. 11] are given in Table I. The sound speeds for the tube walls were initially calculated from the PVC manufacturer's specifications, but the values that were ultimately used (and reported in Table I) were obtained via a calibration procedure that utilized distilled degassed fresh water. With the internal liquid properties and the wall density known, the longitudinal sound speed in the wall was varied in the elastic waveguide model until the predicted phase speeds matched the speed measured during the calibration. There was about a 0.5% difference in phase speeds exhibited by the three resonator tubes when filled with distilled degassed water.

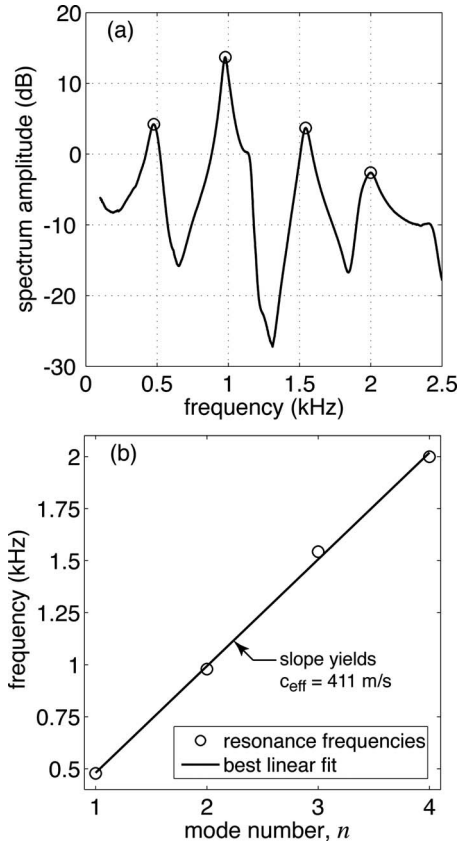


FIG. 2. A typical resonator spectrum (a) for three *Thalassia* leaves. Four resonance peaks are identified. The resulting best-fit sound speed is shown in (b).

V. RESULTS AND DISCUSSION

A typical spectrum from the resonator, obtained with three *Thalassia* leaves, is shown in Fig. 2(a). Four resonance frequencies were identified. The effective sound speed c_{eff} inside the water-and-leaf-filled resonator is inferred from the slope of the curve [Fig. 2(b)], which is the least-squares linear fit to the measured resonance frequencies. The nth resonance frequency is given by $f_n=(c_{\text{eff}}/2L)n$, where L is the length of the water column inside the resonator. Increased leaf biomass inside the resonator (adding additional leaves) shifts the spectrum toward lower resonance frequencies, which indicates a lower effective sound speed and clearly demonstrates that seagrass abundance is directly correlated with an acoustic parameter suitable for remote sensing.

Sound speeds c_{eff} (inferred from the resonance frequencies), free-field sound speeds c_0 (after correction for the elastic waveguide effect), and the apparent void fraction inside the resonator [see Eq. (1)] are reported in Table II. Biomass

TABLE II. Measured resonator sound speeds, associated free-field sound speeds, and apparent resonator void fractions are presented for *Thalassia testudinum*, *Syringodium filiforme*, and *Halodule wrightii*. Also shown are the acoustic- and image-based leaf and rhizome void fractions. The acronym VF is used for "void fraction."

Thalassisa testudinum				
No. of leaves, m	0	3	6	9
Resonator sound speed, c_{eff} (m/s)	422	411	396	393
Free-field sound speed, c_{o} (m/s)	1516	1176	909	874
Apparent acoustic resonator VF, $V_{\text{air}}/V_{\text{tot}}$	0	3.39×10^{-5}	9.12×10^{-5}	1.03×10^{-4}
Acoustically-determined leaf VF	$\chi_{\text{leaf,a}}=0.034$			
Image-based leaf VF	$\chi_{\text{leaf}}=0.23$			
No. of rhizomes, m	0	3	6	
Resonator sound speed, c_{eff} (m/s)	421	409	402	
Free-field sound speed, c_{o} (m/s)	1517	1022	918	
Apparent acoustic resonator VF, $V_{\text{air}}/V_{\text{tot}}$	0	6.13×10^{-5}	8.84×10^{-5}	
Acoustically-determined rhizome VF	$\chi_{\text{rhiz,a}}=0.0047$			
Image-based rhizome VF	$\chi_{\text{rhiz}}=0.13$			
Syringodium filiforme				
No. of leaves, m	0	6	11	25
Resonator sound speed, c_{eff} (m/s)	410	400	396	375
Free-field sound speed, c_{o} (m/s)	1521	1175	1074	791
Apparent acoustic resonator VF, $V_{\text{air}}/V_{\text{tot}}$	0	3.49×10^{-5}	5.17×10^{-5}	1.38×10^{-5}
Acoustically-determined leaf VF	$\chi_{\text{leaf,a}}=0.0038$			
Image-based leaf VF	$\chi_{\text{leaf}}=0.24$			
No. of rhizomes, m	0	1	2	3
Resonator sound speed, c_{eff} (m/s)	408	402	398	393
Free-field sound speed, c_{o} (m/s)	1521	1275	1171	1066
Apparent acoustic resonator VF, $V_{\text{air}}/V_{\text{tot}}$	0	2.21×10^{-5}	3.55×10^{-5}	5.33×10^{-5}
Acoustically-determined rhizome VF	$\chi_{\text{rhiz,a}}=0.0032$			
Image-based rhizome VF	$\chi_{\text{rhiz}}=0.23$			
Halodule wrightii				
No. of leaves, m	0	10	21	33
Resonator sound speed, c_{eff} (m/s)	406	398	390	385
Free-field sound speed, c_{o} (m/s)	1519	1215	1008	946
Apparent acoustic resonator VF, $V_{\text{air}}/V_{\text{tot}}$	0	2.93×10^{-5}	6.55×10^{-5}	8.11×10^{-5}
Acoustically-determined leaf VF	$\chi_{\text{leaf,a}}=0.028$			
Image-based leaf VF	$\chi_{\text{leaf}}=0.19$			
No. of rhizomes, m	0	2	4	5
Resonator sound speed, c_{eff} (m/s)	410.3	407	406	403
Free-field sound speed, c_{o} (m/s)	1517	1358	1347	1244
Apparent acoustic resonator VF, $V_{\text{air}}/V_{\text{tot}}$	0	1.32×10^{-5}	1.42×10^{-5}	2.54×10^{-5}
Acoustically-determined rhizome VF	$\chi_{\text{rhiz,a}}=0.0021$			
Image-based rhizome VF	$\chi_{\text{rhiz}}=0.13$			

as represented by plant tissue volume fraction was also positively correlated with leaf and rhizome number [Figs. 3(a) and 3(b)] in *Thalassia* for both acoustic and image-based measurements. The image-based leaf (χ_{leaf}) and rhizome (χ_{rhiz}) void fractions were determined with the image analysis shown in Figs. 4 and 5. The image-based biomass was compared to the acoustic-based biomass measurement with the following relationship:

$$\left.\frac{V_{\text{leaves}}(m)}{V_{\text{tot}}}\right|_{\text{acoustic}} = \frac{V_{\text{air}}(m)}{V_{\text{tot}}}\frac{1}{\chi_{\text{leaf},a}} = \frac{\chi(m)}{\chi_{\text{leaf},a}}, \tag{2}$$

where $\chi_{\text{leaf},a}$ is the apparent leaf void fraction observed acoustically based on our previous hypotheses and m is the number of leaves. The same approach was applied to rhizome tissue. These calculations reveal a major difference between image- and acoustic-based biomass measurements, es-

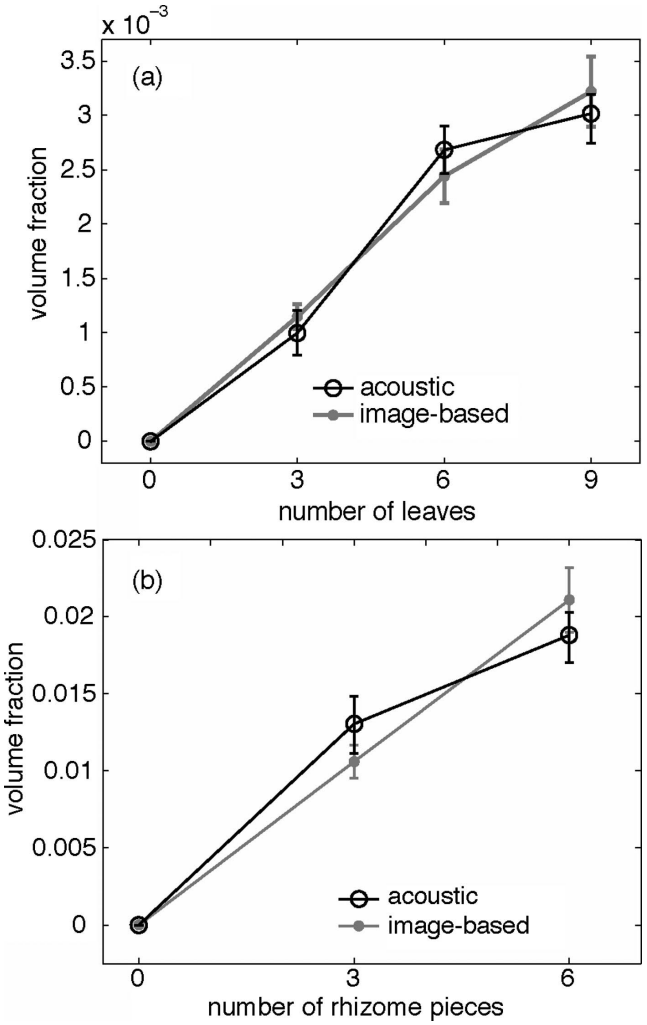


FIG. 3. *Thalassia testudinum.* Comparison of image- and acoustic-based biomass measurements as a function of the number of leaves (a) and rhizomes (b) in the resonator. The vertical axis is total volume fraction V_i/V_{tot} (i=leaves or rhizomes). The error bars are described in the text.

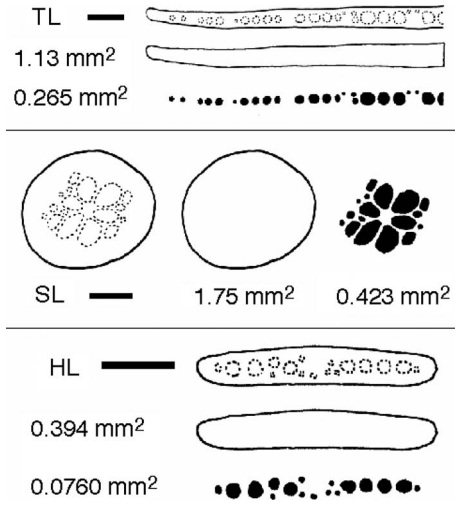


FIG. 4. Microscopic cross-sections of seagrass leaves (L) for *Thalassia* (T; top panel), *Syringodium* (S; middle panel), and *Halodule* (H; bottom panel). Since the cross-sections are nearly uniform perpendicular to the plane of the image, the void fraction is pore area/total area. Leaf sections adapted from Ref. 17. Only half the width of the leaf section is shown for TL. The aerenchyma are the circular features outlined with dashed lines or with black fill. Horizontal black scale bars are 0.5 mm in length. In each panel, the total leaf and pore area are shown adjacent to their corresponding images.

pecially in rhizomes (Fig. 3 and Table II). The best-fit values of $\chi_{\text{leaf},a}$ and $\chi_{\text{rhiz},a}$ (Table II) are means determined from the minimum and maximum values that allowed overlap of the error bars for all the points on both curves (Fig. 3). Such a fitting procedure was used because the curves go through the origin, and hence a linear least-squares approach (fitting only the slope) gave preferential consideration to the final point.

The error bars on the acoustic data represent the effects of uncertainty in the resonator length L and the finite resolution bandwidth of the spectrum measurement. The error bars on the image-based data are due to the finite resolution of the digital images. The leaves and rhizomes of *Thalassia* have extended shapes, where the length is much greater than the thickness. The primary uncertainty is in the accuracy of the thickness measurement, which was limited to about 2 pixels out of 20 in *Thalassi*a. The image-based leaf and rhizome void fractions are 6.8 and 28 times larger than the acoustically-determined void fractions (Table II).

The same procedure was repeated to obtain the physical properties, sound speeds, and void fractions of leaves and rhizomes for *Syringodium filiforme* and *Halodule wrightii* (Tables I and II). The cross-section imagery that yielded tissue void fraction appears in Figs. 4 and 5. The positive relationship between tissue (leaf or rhizome) abundance and volume fraction, as noted in *Thalassia* (Fig. 3), is also readily visible in *Syringodium* (Fig. 6) and *Halodule* (Fig. 7). As in *Thalassia*, however, comparison between the acoustic- and image-based biomass assessments in *Syringodium* and *Halodule* shows significant differences. For all three species, it is clear that biomass is directly proportional to the effective acoustic sound speed of the plant-filled water, but for *Syringodium*, the image-based leaf and rhizome void fractions are 63 and 72 times larger than the acoustically-determined void

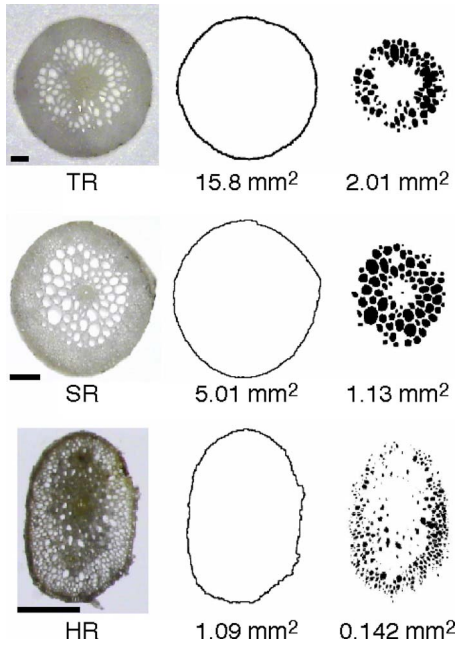


FIG. 5. (Color online) Microscopic cross-sections of rhizomes (R) for the three seagrass species. Abbreviations and notes as in Fig. 4.

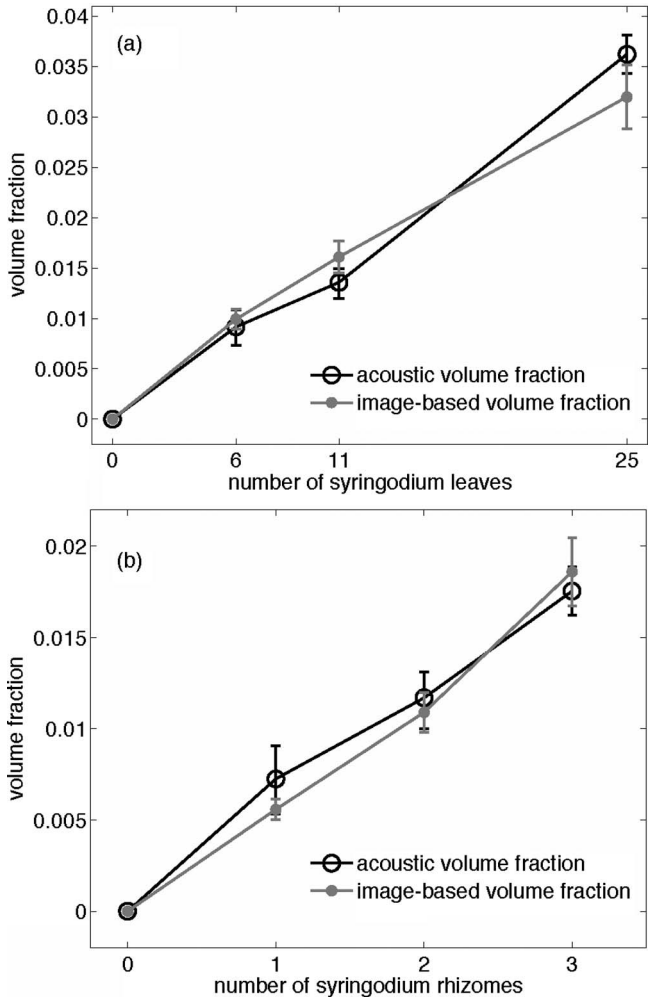


FIG. 6. *Syringodium filiforme*. Comparison of image- and acoustic-based biomass measurements as a function of the number of leaves and rhizomes in the resonator. The vertical axis is total volume fraction V_i/V_{tot} (i =leaves or rhizomes). The error bars are described in the text.

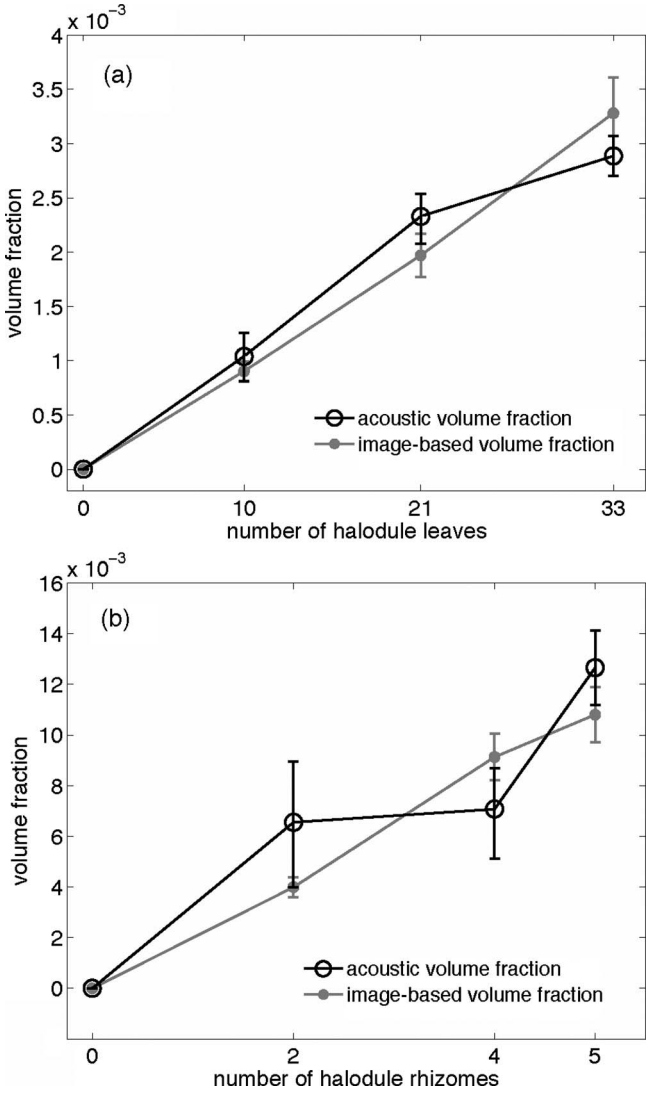


FIG. 7. *Halodule wrightii*. Comparison of image- and acoustic-based biomass measurements as a function of the number of leaves and rhizomes in the resonator. The vertical axis is total volume fraction V_i/V_{tot} (i=leaves or rhizomes). The error bars are described in the text.

fractions (Table II). In *Halodule*, the image-based leaf and rhizome void fractions are 6.8 and 62 times larger than the acoustically-determined void fractions (Table II).

A comparison of the ratio of the image-based to acoustically-determined void fractions for both leaves and rhizomes of all three species reveals some interesting differences (Fig. 8) that reflect the acoustic importance of the tissue. A ratio of unity indicates plants that behave acoustically like air bubbles in water. An increasing ratio indicates increasing tissue stiffness, which effectively reduces the acoustic contrast of the internal gas and thereby reduces the acoustic contrast of the plant. For all three species, the rhizome tissue is stiffer than the leaf tissue, and there is a large difference between the two tissues for both *Thalassia* and *Halodule*. *Syringodium* exhibits the highest tissue stiffness of the three species, and the leaf and tissue stiffness is of similar magnitude. The high stiffness of *Syringodium* leaves may be explained by the cylindrical shape of its above-ground photosynthetic tissues. Volumetric excitation of the pore space places the circumferential tissue in tension, with hoop-like structures resisting expansion. For all three species, the rhizomes are circular in cross-section with internal gas-filled pores, similar to *Syringodium* leaves. In contrast, the leaves of *Thalassia* and *Halodule* are flat and there is little tissue to resist the expansion.

The three species examined here exhibit significantly different acoustic-based and image-based void fractions. Since the image-based void fractions are from direct observations of the plant structure, they were considered to be the best estimate of the actual plant void fraction. One must therefore conclude that hypotheses (1) and (2) are incorrect. Wood's equation is insufficient and the acoustic response of all three species (*Thalassia testudinum*, *Syringodium filiforme*, and *Halodule wrightii*) is dependent on both gas content and the tissue acoustic properties. This corroborates a similar result Hermand[3] found *in situ* for *Posidonia oceanicus.*

There are a number of potential differences between the *in vitro* response observed in the present work and the *in situ* response. First, these experiments were conducted under sufficiently low light conditions to prevent the appearance of bubbles on the leaf surfaces, which indicates that photosyn-

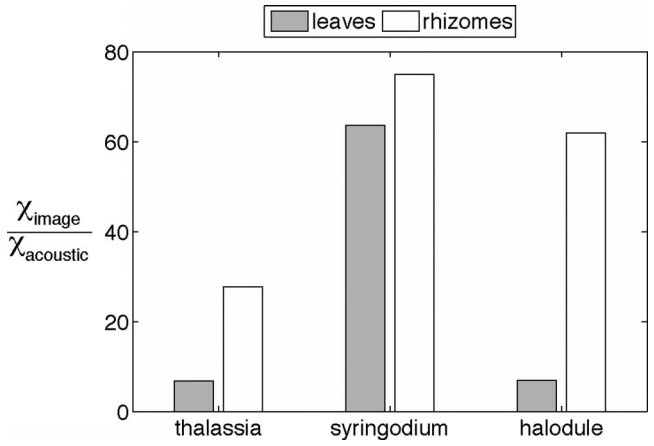


FIG. 8. The ratio of the image-based to acoustically-determined void fractions for both leaves and rhizomes of each species. This ratio is a measure of the importance of the tissue acoustic parameters. A ratio of unity indicates plants that behave acoustically like air bubbles in water. An increasing ratio indicates increasing tissue stiffness, which effectively reduces the acoustic contrast of the internal gas and also reduces the acoustic contrast of the plant.

thetic respiration was not underway during the experiments. *In situ*, under different light conditions and hence, different photosynthetic respiration conditions, the acoustic response of all three species will likely be different because the gas content and internal pressure of the plant will be different. A limited number of measurements under light saturation conditions were conducted and significant bubble production on the leaves of *Thalassia* was observed, which in turn further reduced the effective resonator sound speed, as expected. This was accompanied by a significant temperature increase in the resonator to unnaturally high levels. Resonator temperature control will have to be implemented in the present apparatus to accurately investigate the effect of illumination level. Second, the results reported here are for near-surface conditions. Because the sound speed in a gas is dependent on the ambient pressure, the acoustic response of seagrass will also be a function of depth. Third, during the present experiments, the acoustic axis was oriented along the length of the leaves whereas *in situ*, the acoustic axis could be oriented perpendicular to the leaves. Since some seagrass species have long, thin, relatively flat leaves, there is likely an excitation frequency above which the effective acoustic properties for cross-leaf propagation may be a function of leaf orientation. Despite these complexities, this work shows that seagrass biomass is directly related to the low frequency effective sound speed in a body of water containing seagrass plants. Further, the ability to remotely detect gas content in both above- and below-ground tissues may provide a reliable technique to rapidly assess photosynthetic production in seagrass beds that are chronically exposed to limiting light conditions in estuarine environments.[3,7,14]

VI. CONCLUSIONS

The low frequency acoustic properties of seagrass leaves and rhizomes were investigated *in vitro* using an acoustic resonator technique and image-based biomass measurements. The effective sound speed inside a plant-and-water-filled resonator was found to be proportional to overall tissue biomass contained in the resonator. The effective sound speed decreased as the number of leaves (or rhizome pieces) increased. The change in sound speed was reduced for physically smaller leaf species (*Halodule* < *Syringodium* < *Thalassia*). Gas content is clearly important, but since gas content (via Wood's equation) by itself did not describe the observed sound speeds, the acoustic response of the tissue must also be considered in the development of future acoustic models of sound propagation in seagrass beds. Finally, we found that flat leaves exhibited more acoustic contrast than did circular cross-section leaves or rhizomes, potentially due to hoop stress in the outer layers of the circular cross-section tissues characteristic of *Syringodium*.

The present work is a first step toward modeling sound propagation in seagrass beds. Additional work is needed to fully understand even the simplified laboratory experiments reported here. Significant additional work is needed to understand the complexities mentioned above and additional complexity found *in situ*, such as diurnal and seasonal variations.

ACKNOWLEDGMENTS

We thank K. Jackson for providing fresh samples of field-collected seagrasses and Chad Greene for assistance with the image analysis. Work supported by the Office of Naval Research Ocean Acoustics Program, the UT College of Engineering, and the Texas Sea Grant College Program Grant No. R/ES-87.

[1]E. M. McCarthy and B. Sabol, "Acoustic characterization of submerged aquatic vegetation: Military and environmental monitoring applications," in *Proceedings of OCEANS 2000 MTS/IEEE Conference and Exhibition* (IEEE, Piscataway, NJ, 2000), Vol. **3**, pp. 1957–1961.
[2]A. P. Lyons and D. A. Abraham, "Statistical characterization of high-frequency shallow-water seafloor backscatter," J. Acoust. Soc. Am. **106**, 1307–1315 (1999).
[3]J. P. Hermand, in *Handbook of Scaling Methods in Aquatic Ecology: Measurement, Analysis, Simulation*, edited by L. Seuront and P. G. Strutton (CRC, Boca Raton, FL, 2004), pp. 65–96.
[4]V. Pasqualini, P. Clabaut, G. Pergent, L. Benyoussef, and C. Pergent-Martini, "Contribution of side scan sonar to the management of Mediterranean littoral ecosystems," Int. J. Remote Sens. **21**, 367–378 (2000).
[5]P. J. Mulhearn, "Mapping seabed vegetation with sidescan sonar," DSTO Report No. TN-0381, Defence Science and Technology Organization, Australia, 2001.
[6]T. Komatsu, C. Igarashi, K. Tatsukawa, S. Sultana, Y. Matsuoka, and S. Harada, "Use of multi-beam sonar to map seagrass beds in Otsuchi Bay on the Sanriku Coast of Japan," Aquat. Living Resour. **16**, 223–230 (2003).
[7]J. P. Hermand, P. Nascetti, and F. Cinelli, "Inversion of acoustic waveguide propagation features to measure oxygen synthesis by Posidonia oceanica," in *OCEANS '98 Conference Proceedings* (IEEE, Piscataway, NJ, 1998), Vol. **2**, pp. 919–926.
[8]A. B. Wood, *A Textbook of Sound*, 1st ed. (MacMillan, New York, 1930).
[9]R. C. Phillips and E. G. Meñez, *Seagrasses* (Smithsonian Institution Press, Washington, DC, 1988).
[10]P. S. Wilson, "Low-frequency dispersion in bubbly liquids," ARLO **6**, 188–194 (2005).
[11]P. S. Wilson, A. H. Reed, J. C. Wilbur, and R. A. Roy, "Evidence of dispersion in an artificial water-saturated sand sediment," J. Acoust. Soc. Am. **121**, 824–832 (2007).
[12]P. S. Wilson, R. A. Roy, and W. M. Carey, "Phase speed and attenuation in bubbly liquids inferred from impedance measurements near the individual bubble resonance frequency," J. Acoust. Soc. Am. **117**, 1895–1910 (2005).
[13]J. S. Bendat and A. G. Piersol, *Engineering Applications of Correlation and Spectral Analysis*, 2nd ed. (Wiley, New York, 1993).

[14]K. H. Dunton, "Seasonal growth and biomass of the subtropical seagrass *Halodule wrightii* in relation to continuous measurements of underwater irradiance," Mar. Biol. (Berlin) **120**, 479–489 (1994).
[15]N. P. Fofonoff, "Physical properties of seawater: A new salinity scale and equation of state for seawater," J. Geophys. Res. **90**, 3332–3342 (1985).
[16]National Oceanographic Data Center, "Coastal water temperature guide–Gulf of Mexico Coast: Western," http://www.nodc.noaa.gov/dsdt/cwtg/wgof.html (Last viewed April, 2008).
[17]R. C. Phillips and C. P., McRoy, *Handbook of Seagrass Biology: An Ecosystem Perspective* (Garland STPM, New York, 1980).
[18]V. A. Del 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).

Retrieval of Green's functions of elastic waves from thermal fluctuations of fluid-solid systems

Oleg A. Godin[a)]
CIRES, University of Colorado, 216 UCB, Boulder, Colorado 80309-0216 and NOAA/Earth System Research Laboratory, DSRC, Mail Code R/PSD99, 325 Broadway, Boulder, Colorado 80305-3328

(Received 21 October 2008; accepted 22 January 2009)

Fluctuation-dissipation and flow reversal theorems are used to study long-range correlation of thermal phonons in a stationary heterogeneous mechanical system comprised of arbitrary inhomogeneous fluid flow and anisotropic solid. At thermal equilibrium, with an appropriate choice of physical observables to characterize thermal fluctuations within the fluid and within the solid, the general integral expression for the two-point correlation function of the fluctuations reduces to a linear combination of deterministic Green's functions, which describe wave propagation in opposite directions between the two points. It is demonstrated that the cross-correlation of thermal noise contains as much information about the environment as can be obtained in active reciprocal transmission experiments with transceivers placed at the two points. These findings suggest a possible application of ambient noise cross-correlation to passive acoustic characterization of inhomogeneous flows in fluid-solid systems in laboratory and geophysical settings.
 [DOI: 10.1121/1.3082101]

PACS number(s): 43.30.Pc, 43.20.Bi, 43.40.Le, 43.60.Rw [RLW] Pages: 1960–1970

I. INTRODUCTION

Rather counterintuitively, linear wave fields possess significant correlations at ranges that are large compared to the wavelength, even when the waves are generated by random delta-correlated sources. Two-point correlation function of diffuse wave fields is closely related to the Green's function describing wave propagation between the two points. Apparently, Rytov was the first to derive a simple algebraic relation[1,2] between the spatial cross-correlation function of diffuse noise and the deterministic Green's function. Rytov considered electromagnetic fields of thermal origin in motionless dielectrics and applied his theory of fluctuational electrodynamics to extend Kirchhoff's law of thermal radiation, calculate thermal noise in extended receiving arrays, and develop a theory of macroscopic cohesion forces.[2,3] Much later, Rytov's relation was rediscovered in the seismology community.[4–6] A broad interest in the relation between long-range noise correlation and characteristics of deterministic wave propagation was generated by seminal experiments of Weaver and Lobkis.[7–9] They demonstrated that cross-correlation of thermal phonons[7,8] and other diffuse fields of elastic waves[9] in solids received at two points gives the time-domain Green's function between the points, i.e., the wave field that would be observed at one point if an impulsive point source were placed at the other. Green's function contains as complete environmental information as can be obtained from wave transmission between the two points. Thus, probing of an environment by emitting and receiving waves can be replaced, in principle, by cross-correlation of ambient noise.

Ambient noise tomography has rapidly become an extremely valuable technique in seismology[10–20] and helioseismology.[21–23] Promising applications of ambient noise cross-correlation to passive remote sensing have been proposed in underwater acoustics,[24–29] mechanical and civil engineering,[30–33] and medical acoustics.[34,35]

Theoretically, the relation between cross-correlation of random fields in a stationary medium and the deterministic acoustic Green's function has been established using normal modes of a closed system,[7–9] time-reversal arguments,[36,37] stationary phase analysis,[26,38–40] or reciprocity considerations;[6,40–43] for recent reviews see Weaver and Lobkis[44] and Gouédard *et al.*[16] Extension of this work to arbitrary inhomogeneous moving fluids showed that the cross-correlation of an acoustic field due to random sources distributed on a surface gives a combination of acoustic Green's functions, which describe sound propagation in opposite directions between the two points, and contains information necessary to characterize both sound speed and flow velocity.[40] Similar results, but for a fluid with slow and weakly varying currents, were independently obtained in Refs. 42 and 45. Recently, it was shown that cross-correlation of thermal phonons in an arbitrary inhomogeneous fluid flow in thermal equilibrium also gives a combination of Green's functions corresponding to sound propagation in opposite directions.[43] These results suggest the possibility of passive remote sensing of inhomogeneous flows, including precise measurements of flow velocities that are small compared to sound speed, which is highly relevant to environmental monitoring and medical diagnostics.

In the laboratory as well as in geophysical settings, elastic waves are normally encountered in environments comprising both fluid and solid components, although the effect of the fluid on waves within a solid and vice versa is sometimes negligible. In other problems, such as wave propagation in porous media and passive acoustic monitoring of either ocean dynamics with sensors placed on the ocean

[a)]Electronic mail: oleg.godin@noaa.gov

bottom or a fluid motion within a vessel with sensors located on its external wall, an account of coupling between waves in the fluid and the solid is paramount.[46] Acoustic noise in fluid-solid environments with flow can have diverse origins, including cavitation noise, aerodynamic noise, or biological and shipping noise in the ocean. Here we consider thermal noise, i.e., waves representing chaotic thermal motion of molecules, atoms, ions, and electrons in a coupled fluid-solid system. Unlike other types of acoustic noise, thermal phonons are inevitably present in fluids and solids. Thermal noise often dominates the noise field at sufficiently high frequencies.[7,8,34,47–52] Measurements of intensity and correlation properties of thermal radiation underlie acoustic brightness thermometry and thermoacoustic tomography, which have important biomedical applications.[48,51,53,54]

In this paper, we apply the fluctuation-dissipation[3,55] and flow reversal[56–58] theorems (FRTs) to calculate statistical moments of thermal fluctuations in rather general fluid-solid systems, which comprise arbitrary inhomogeneous fluid flow and an anisotropic solid. We show that time-domain Green's functions corresponding to wave propagation in both directions between any two points can be found as time derivatives of the cross-correlation function of thermal noise measured at the two points.

The plan of the paper is as follows. Equations governing wave fields in moving fluids and anisotropic solids as well as boundary conditions at boundaries and interfaces are formulated in Sec. II. The FRT, which expresses reciprocity of wave fields in fluid-solid systems, is discussed in Sec. III. In Sec. IV, the theorem is applied (i) to derive general expressions for wave fields generated by extraneous forces and strains and (ii) to establish certain integral identities for the Green's functions. Effective macroscopic sources of thermal noise are found in Sec. V using a spectral formulation[3] of the fluctuation-dissipation theorem (FDT). Results of Secs. IV and V are combined in Sec. VI to calculate the cross-correlation function of thermal fluctuations in terms of deterministic Green's functions of elastic waves, which correspond to various types of system excitation. Implications for passive remote sensing and relation to earlier work are addressed in Sec. VII.

II. GOVERNING EQUATIONS AND BOUNDARY CONDITIONS

Consider elastic waves in a stationary (i.e., independent of time t) heterogeneous environment consisting of a compressible fluid moving past a deformable solid structure (Fig. 1). Linear acoustic fields in an inhomogeneous moving fluid with sound speed $c(\mathbf{x})$, mass density $\rho(\mathbf{x})$, and flow velocity $\mathbf{U}(\mathbf{x})$ are governed by the equations[56,57]

$$\rho\frac{d^2\mathbf{w}}{dt^2}+\nabla p+(\mathbf{w}\cdot\nabla)\,\nabla\, p_0-\frac{\nabla p_0}{\rho c^2}(p+\mathbf{w}\cdot\nabla p_0)=\mathbf{F}, \quad (1)$$

$$\nabla\cdot\mathbf{w}+(p+\mathbf{w}\cdot\nabla p_0)/\rho c^2=B, \quad d/dt\equiv\partial/\partial t+\mathbf{U}\cdot\nabla, \quad (2)$$

where p, p_0, and $\mathbf{w}$ are acoustic pressure, background pressure (i.e., the pressure in the absence of acoustic waves), and oscillatory displacement of fluid particles due to an acoustic wave. The acoustic wave is generated by sources with volume densities $\mathbf{F}$ of force and dB/dt of volume velocity. The gradient of the background pressure is given by the Euler equation as $\nabla p_0=-\rho(\mathbf{U}\cdot\nabla)\mathbf{U}$. The oscillatory displacement $\mathbf{w}$ is related to the oscillatory velocity of fluid particles, i.e., the linear perturbation of fluid velocity due to acoustic wave, by the equation[56,57] $\mathbf{v}=d\mathbf{w}/dt-(\mathbf{w}\cdot\nabla)\mathbf{U}$. For continuous waves (CWs) we will assume and suppress the time dependence $\exp(-i\omega t)$.

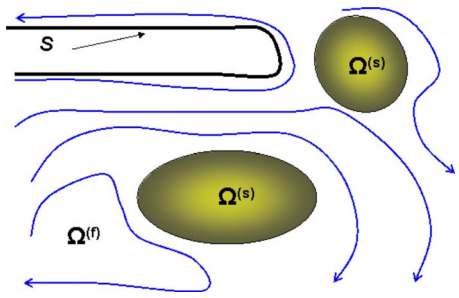


FIG. 1. (Color online) Geometry of the problem. A heterogeneous mechanical system is comprised of moving fluid, which occupies domain $\Omega^{(f)}$, and anisotropic solid, which occupies domain $\Omega^{(s)}$. Aside from fluid-solid interfaces, the boundary $\partial\Omega^{(f)}$ of the fluid may contain pressure-release, rigid, or impedance surfaces S.

Equations (1) and (2) have been derived[56,57] neglecting irreversible thermodynamic processes. It is common to model sound propagation in a medium with dissipation by formally attributing frequency-dependent complex values to sound speed and mass density.[59] With complex $c(\mathbf{x},\omega)$ and $\rho(\mathbf{x},\omega)$, Eqs. (1) and (2) become model governing equations for CW acoustic fields in a moving fluid with dissipation. Note that assigning complex values to sound speed and mass density results in only an approximate description of the physical processes in the fluid with dissipation. In particular, the model does not account for viscous and thermal waves (see Brekhovskikh and Godin[59]). These rapidly attenuating waves are not important for long-range field correlations, which are the subject of this paper.

The linearized boundary conditions[56,57] for the CW acoustic field at an impedance (i.e., locally reacting) surface S within a moving fluid are

$$p+(\mathbf{N}\cdot\nabla p_0-i\omega\zeta)\mathbf{w}\cdot\mathbf{N}=0, \quad \mathbf{x}\in S, \quad (3)$$

where $\mathbf{N}$ is a unit normal to S and $\zeta=\zeta(\mathbf{x},\omega)$ is the impedance of the surface. The impedance is a property of the surface and is supposed to be independent of the background flow velocity. Free and rigid boundaries can be viewed as special cases of the impedance surface corresponding to zero or infinite impedance, respectively. In addition, the normal component of the background flow velocity is zero[57] on S and any boundary impenetrable to fluid:

$$\mathbf{U}\cdot\mathbf{N}=0. \quad (4)$$

In a quiescent fluid ($\mathbf{U}\equiv 0$), Eqs. (1) and (2) reduce to standard equations,[57,59]

$$\rho\frac{\partial \mathbf{v}}{\partial t}+\nabla p=\mathbf{F},\quad \nabla\cdot\mathbf{v}+\frac{1}{\rho c^2}\frac{\partial p}{\partial t}=\frac{\partial B}{\partial t},\tag{5}$$

of linear acoustics of motionless media. Some authors[60] do not appreciate the well-known fact that, unless $\mathbf{U}=\text{const}$, replacing partial time derivatives[42,45,60] $\partial/\partial t$ with material time derivatives d/dt in Eq. (5) does not lead to the correct governing equations for sound in an inhomogeneous flow. Strictly speaking, use of such model equations instead of the correct governing equations for sound in moving fluid in Refs. 42, 45, and 60 limits the validity domain of their conclusions to the trivial case of uniform flow.

Linear waves in an arbitrary inhomogeneous and anisotropic solid are governed by the equations[59,61]

$$\rho\partial^2 w_j/\partial t^2=\partial\sigma_{jk}/\partial x_k+F_j,\tag{6}$$

$$\sigma_{jk}=C_{jkmn}(u_{mn}+e_{mn}),\quad u_{mn}=\frac{1}{2}\left(\frac{\partial w_m}{\partial x_n}+\frac{\partial w_n}{\partial x_m}\right),\tag{7}$$

where $\mathbf{w}$, u_{jk}, and σ_{jk} are, respectively, displacement of particles and strain and stress tensors in the wave generated by an external force with volume density $\mathbf{F}$ and an external strain e_{mn}; C_{jkmn} is the elastic modulus tensor. Here and below summation over repeated Latin subscripts j, k, m, n, l, $s=1,2,3$ is implied. In terms of the compliance tensor α_{jkmn}, the stress-strain relationship (7) can be alternatively written as

$$\alpha_{mnjk}\sigma_{jk}=u_{mn}+e_{mn}.\tag{8}$$

The tensors u_{jk}, e_{mn}, and σ_{jk} are symmetric: $u_{jk}=u_{kj}$, $e_{mn}=e_{nm}$, $\sigma_{jk}=\sigma_{kj}$, while the tensors C_{jkmn} and α_{jkmn} are invariant with respect to interchange of the indices within the first or second pair and of the pairs. These symmetry properties are systematically utilized in the transformations in Secs. III–VI. As long as σ_{jk} and $u_{nm}+e_{mn}$ can be arbitrary symmetric tensors, the compliance and elastic modulus tensors are related by the equation

$$\alpha_{mnjk}C_{jkrs}=\tfrac{1}{2}(\delta_{mr}\delta_{ns}+\delta_{nr}\delta_{ms}),\tag{9}$$

where δ_{mn} is the Kronecker delta symbol. Dissipation of CW elastic waves in a solid is modeled[59,61] by attributing complex values to density and material tensors C_{jkmn} and α_{jkmn}. As in the case of fluid, we disregard rapidly attenuating thermal waves in a solid. Below we will need the following relation between imaginary parts of the compliance and elastic modulus tensors:

$$C_{jkrs}C^{*}_{mnlq}\,\mathrm{Im}\,\alpha_{mnjk}=-\,\mathrm{Im}\,C_{lqrs},\tag{10}$$

which is a corollary of the identity (9). Here and below, the asterisk denotes complex conjugation.

Equations (6) and (7) for elastic waves in solids contain a description of acoustic waves in a motionless fluid (i.e., in the absence of a background flow). When $C_{jkmn}=\rho c^2\delta_{jk}\delta_{mn}$ and $\sigma_{jk}=-p\delta_{jk}$ Eqs. (6) and (7) reduce to Eqs. (1) and (2) with $\mathbf{U}=0$ and become equivalent to Eq. (5). Note that sound in fluid is generated by a source of external strain only when resulting volume change is nonzero: $e_{nn}\neq 0$. Then the source of external strain is equivalent to a source of volume velocity with $B=-e_{nn}/3$. These relations are utilized below to verify results obtained for waves in fluid-solid systems against known results for acoustic fields in fluids.

Boundary conditions on rigid and free boundaries of a solid are[59] $\mathbf{w}=0$ and $\sigma_{jk}N_k=0$, respectively. Linearized kinematic boundary condition[57,58] on a fluid-solid interface Γ consists in continuity of the normal displacement $\mathbf{w}\cdot\mathbf{N}$. Generally, a solid in contact with inhomogeneous fluid flow is deformed even in the background state, i.e., in the absence of waves. Assuming that the background stress tensor $\sigma^{(0)}_{jk}$ and its normal derivative are diagonal on the interface, the linearized dynamic boundary condition[58] can be written as

$$\sigma_{jk}N_j+[p+(\mathbf{w}\cdot\mathbf{N})H_0]N_k=0,\quad \mathbf{x}\in\Gamma,\tag{11}$$

where

$$H_0=\mathbf{N}\cdot\nabla p_0+\tfrac{1}{3}(\mathbf{N}\cdot\nabla)\sigma^{(0)}_{nn}.\tag{12}$$

III. RECIPROCITY RELATIONS IN FLUID-SOLID SYSTEMS

The FRT[56–58] relates the wave field $p^{(1)},\mathbf{w}^{(1)},\sigma^{(1)}_{jk}$ generated by a set $B^{(1)},\mathbf{F}^{(1)},e^{(1)}_{jk}$ of sources in a medium with flow velocity $\mathbf{U}(\mathbf{x})$, with an acoustic field $p^{(2)},\mathbf{w}^{(2)},\sigma^{(2)}_{jk}$ generated by another set $B^{(2)},\mathbf{F}^{(2)},e^{(2)}_{jk}$ of sources in a medium with the same geometry of the interfaces and boundaries, the same sound speed, elastic modulus tensor and density, and *reversed* flow with velocity $-\mathbf{U}(\mathbf{x})$. Note that the flow reversal does not change the background pressure field $p_0(\mathbf{x})$. Using Eqs. (1) and (2) to calculate the quantity

$$a^{(f)}=\mathbf{F}^{(2)}\mathbf{w}^{(1)}-\mathbf{F}^{(1)}\mathbf{w}^{(2)}+B^{(1)}p^{(2)}-B^{(2)}p^{(1)}\tag{13}$$

and Eqs. (6) and (7) to calculate the quantity

$$a^{(s)}=\mathbf{F}^{(2)}\mathbf{w}^{(1)}-\mathbf{F}^{(1)}\mathbf{w}^{(2)}+e^{(1)}_{mn}\sigma^{(2)}_{mn}-e^{(2)}_{mn}\sigma^{(1)}_{mn},\tag{14}$$

and applying the Gauss divergence theorem, we obtain

$$\int_{\Omega^{(f)}}a^{(f)}d^3\mathbf{x}+\int_{\Omega^{(s)}}a^{(s)}d^3\mathbf{x}=\int_{\partial\Omega^{(f)}}\mathbf{N}\cdot\mathbf{j}^{(f)}ds+\int_{\partial\Omega^{(s)}}\mathbf{N}\cdot\mathbf{j}^{(s)}ds.\tag{15}$$

Here $\Omega^{(f)}$ and $\Omega^{(s)}$ are arbitrary domains occupied by fluid and solid, respectively, with the boundaries $\partial\Omega^{(f)}$ and $\partial\Omega^{(s)}$, $\mathbf{N}$ is an external unit normal to the boundaries, and

$$\mathbf{j}^{(f)}=p^{(2)}\mathbf{w}^{(1)}-p^{(1)}\mathbf{w}^{(2)}-\rho\mathbf{U}(\mathbf{w}^{(1)}\cdot\tilde{d}\mathbf{w}^{(2)}/dt+\mathbf{w}^{(2)}\cdot d\mathbf{w}^{(1)}/dt),\tag{16}$$

$$j^{(s)}_k=\sigma^{(1)}_{jk}w^{(2)}_j-\sigma^{(2)}_{jk}w^{(1)}_j.\tag{17}$$

The tilde denotes quantities referring to a medium with reversed flow. In particular, $\tilde{d}/dt=\partial/\partial t-\mathbf{U}\cdot\nabla$. The boundaries $\partial\Omega^{(f)}$ and $\partial\Omega^{(s)}$ are not required to be simply connected. Parts of $\partial\Omega^{(f)}$ and $\partial\Omega^{(s)}$ may be located at infinity. Derivation of the FRT (13)–(17) parallels that presented in Ref. 58 and differs only by inclusion of the external strain as additional type of wave sources in the solid.

Below we assume that, aside from fluid-solid interfaces, the fluid is either unbounded or has ideal (i.e., free or rigid) or impedance boundaries, while the solid is either unbounded or has ideal boundaries. According to Eqs. (3), (4), (16), and (17), normal components of the vectors $\mathbf{j}^{(f)}$ and $\mathbf{j}^{(s)}$ vanish on such boundaries. According to Eqs. (4), (11), (16), and (17), $(\mathbf{j}^{(f)}-\mathbf{j}^{(s)})\cdot\mathbf{N}=0$ at the fluid-solid interface. As external normals to the surfaces $\partial\Omega^{(f)}$ and $\partial\Omega^{(s)}$ have opposite directions on the fluid-solid interface, contributions of such interfaces into integrals over $\partial\Omega^{(f)}$ and over $\partial\Omega^{(s)}$ in Eq. (15) cancel each other. Parts of $\partial\Omega^{(f)}$ and $\partial\Omega^{(s)}$ that are located at infinity do not contribute to the surface integrals in Eq. (15).[56–58] Thus, when the domain $\Omega\equiv\Omega^{(f)}\cup\Omega^{(s)}$ is chosen to be the entire volume occupied by the fluid and the solid, the right side of Eq. (15) vanishes.

In the absence of a background flow ($\mathbf{U}\equiv 0$), FRT (13)–(17) relates wave fields created by two different sets of sources in the same environment and extends the acoustic reciprocity principle[57] and the Betti reciprocal relation of dynamic elasticity[61] to fluid-solid systems.

The FRT relates wave fields in the original medium and in the medium with reversed flow. A similar relation can be derived for a wave field due to a set of sources and a complex conjugate of a wave field due to another set of sources in the same medium. Using Eqs. (1) and (2) to calculate the quantity

$$A^{(f)}=\mathbf{F}^{(2)*}\mathbf{w}^{(1)}-\mathbf{F}^{(1)}\mathbf{w}^{(2)*}+B^{(1)}p^{(2)*}-B^{(2)*}p^{(1)} \tag{18}$$

and Eqs. (6) and (7) to calculate the quantity

$$A^{(s)}=\mathbf{F}^{(2)*}\mathbf{w}^{(1)}-\mathbf{F}^{(1)}\mathbf{w}^{(2)*}+e^{(1)}_{mn}C_{jkmn}\,\partial\, w^{(2)*}_{j}/\partial x_k - e^{(2)*}_{mn}C^{*}_{jkmn}\,\partial\, w^{(1)}_{j}/\partial x_k, \tag{19}$$

after application of the Gauss divergence theorem we obtain

$$\begin{aligned}\int_{\Omega^{(f)}}A^{(f)}d^3\mathbf{x}+\int_{\Omega^{(s)}}A^{(s)}d^3\mathbf{x}&=\int_{\partial\Omega^{(f)}}\mathbf{N}\cdot\mathbf{J}^{(f)}ds+\int_{\partial\Omega^{(s)}}\mathbf{N}\cdot\mathbf{J}^{(s)}ds+2i\int_{\Omega^{(f)}}\left(\mathrm{Im}\frac{1}{\rho c^2}\right)\\&\times(p^{(1)}+\mathbf{w}^{(1)}\cdot\nabla p_0)(p^{(2)}+\mathbf{w}^{(2)}\cdot\nabla p_0)^{*}d^3\mathbf{x}\\&+2i\int_{\Omega^{(s)}}\left[\omega^2\mathbf{w}^{(1)}\cdot\mathbf{w}^{(2)*}\,\mathrm{Im}\,\rho-\frac{\partial w^{(1)}_m}{\partial x_n}\frac{\partial w^{(2)*}_j}{\partial x_k}\mathrm{Im}\,C_{jkmn}\right]d^3\mathbf{x}.\end{aligned} \tag{20}$$

Here

$$\mathbf{J}^{(f)}=p^{(2)*}\mathbf{w}^{(1)}-p^{(1)}\mathbf{w}^{(2)*}+\rho\mathbf{U}[\mathbf{w}^{(1)}(\mathbf{U}\cdot\nabla)\mathbf{w}^{(2)*}-\mathbf{w}^{(2)*}(\mathbf{U}\cdot\nabla)\mathbf{w}^{(1)}+2i\omega\mathbf{w}^{(1)}\cdot\mathbf{w}^{(2)*}], \tag{21}$$

$$J^{(s)}_k=\sigma^{(1)}_{jk}w^{(2)*}_j-\sigma^{(2)*}_{jk}w^{(1)}_j, \tag{22}$$

and $\Omega^{(f)}$, $\Omega^{(s)}$, $\partial\Omega^{(f)}$, and $\partial\Omega^{(s)}$ have the same meaning as in Eq. (15). In derivation of Eq. (20), it has been assumed that acoustic field dissipation in fluid is described by an imaginary part of the fluid compressibility $1/\rho c^2$, while the other fluid parameters are real-valued. Note that, according to Eqs. (21) and (22), those parts of the surfaces $\partial\Omega^{(f)}$ and $\partial\Omega^{(s)}$, which are located at infinity or coincide with ideal boundaries, do not contribute to surface integrals in the right-hand side of Eq. (20), as in the FRT (15). According to the boundary condition (3), the normal component of the vector $\mathbf{J}^{(f)}$ [Eq. (21)] is also zero on an impedance boundary provided that the impedance is purely reactive ($\mathrm{Re}\,\zeta=0$). From the boundary conditions (4) and (11) and continuity of $\mathbf{w}\cdot\mathbf{N}$, it follows that fluxes of vectors $\mathbf{J}^{(f)}$ [Eq. (21)] and $\mathbf{J}^{(s)}$ [Eq. (22)] through a fluid-solid interface cancel each other. Thus, when the domain $\Omega\equiv\Omega^{(f)}\cup\Omega^{(s)}$ is chosen to be the entire volume occupied by the fluid and the solid and the medium is either unbounded or has only ideal boundaries or purely reactive impedance boundaries, the surface integrals can be omitted in the right-hand side of Eq. (20).

The identity (20) extends to flow-structure problems the "correlation-type" reciprocity relations, which have been previously derived for waves in inhomogeneous solids[42] and moving fluids[43] with dissipation as well as in flow-structure problems without dissipation.[58] In the particular case[57,58] where the wave field with the superscript (2) is chosen to be the same as the wave field with the superscript (1), Eq. (20) expresses conservation of wave energy, with the imaginary part of the surface integrals in the right-hand side giving time-averaged energy flux through the boundaries of the fluid and solid domains $\Omega^{(f)}$ and $\Omega^{(s)}$, volume integrals giving wave energy dissipation in the fluid and the solid, and the imaginary part of the left-hand side describing time-averaged power output of the wave sources.

IV. GREEN'S FUNCTIONS

As in Refs. 40 and 43, we define the Green's function $G(\mathbf{x},\mathbf{x}_1,t)$ as the acoustic pressure at point $\mathbf{x}$ in fluid due to a point source of volume velocity with $B=\delta(\mathbf{x}-\mathbf{x}_1)\delta'(t)$ and $\mathbf{F}=0$. The stress tensor σ_{jk} in the solid and oscillatory displacement $\mathbf{w}$ generated by such an acoustic source will be denoted as $s_{jk}(\mathbf{x},\mathbf{x}_1,t)$ and $\mathbf{g}(\mathbf{x},\mathbf{x}_1,t)$, respectively. In the case of a motionless fluid, the Green's function so defined corresponds to the field of a monopole sound source. The frequency spectrum $G(\mathbf{x},\mathbf{x}_1,\omega)$ of the time-domain Green's function $G(\mathbf{x},\mathbf{x}_1,t)$ has the meaning of the CW Green's function. In a motionless fluid, $\mathbf{g}(\mathbf{x},\mathbf{x}_1,\omega)=(\rho\omega^2)^{-1}\nabla G(\mathbf{x},\mathbf{x}_1,\omega)$. For the Green's functions G and $\tilde{G}$ in the original medium and in a medium with reversed flow, a simple symmetry relation follows from Eq. (15):

$$G(\mathbf{x},\mathbf{x}_1,\omega)=\tilde{G}(\mathbf{x}_1,\mathbf{x},\omega). \tag{23}$$

From the Green's function definition and FRT (15), we obtain

$$p(\mathbf{x},\omega)=\frac{i}{\omega}\int d^3\mathbf{y}[B(\mathbf{y})\tilde{G}(\mathbf{y},\mathbf{x},\omega)-\mathbf{F}(\mathbf{y})\cdot\tilde{\mathbf{g}}(\mathbf{y},\mathbf{x},\omega)+e_{mn}(\mathbf{y})\tilde{s}_{mn}(\mathbf{y},\mathbf{x},\omega)] \tag{24}$$

for the acoustic field in the fluid, which is jointly generated by the distributed CW sources of an external force with an arbitrary density $\mathbf{F}$, of volume velocity with a density B within the fluid, and of an external strain with a density e_{mn}

in the solid. The integral in Eq. (24) is taken over the entire domain occupied by the sources.

To derive another important identity for the Green's function $G(\mathbf{x},\mathbf{x}_1,\omega)$, we apply the reciprocity relation (20) to wave fields generated by sources with the frequency spectra $B^{(r)}=-i\omega\delta(\mathbf{x}-\mathbf{x}_r)$, $\mathbf{F}^{(r)}=0$, and $e_{mn}^{(r)}=0$, where $r=1,2$. With $\Omega\equiv\Omega^{(f)}\cup\Omega^{(s)}$ representing the whole domain occupied by the fluid and the solid and assuming that medium boundaries, if any, are free, rigid, or locally reacting surfaces with purely reactive impedance, we obtain

$$
\begin{aligned}
&-0.5\omega[G(\mathbf{x}_2,\mathbf{x}_1,\omega)+G^*(\mathbf{x}_1,\mathbf{x}_2,\omega)]\\
&=\int_{\Omega^{(f)}}\left(\mathrm{Im}\frac{1}{\rho c^2}\right)[G(\mathbf{x},\mathbf{x}_1,\omega)+\mathbf{g}(\mathbf{x},\mathbf{x}_1,\omega)\cdot\nabla p_0]\\
&\quad\times[G(\mathbf{x},\mathbf{x}_2,\omega)+\mathbf{g}(\mathbf{x},\mathbf{x}_2,\omega)\cdot\nabla p_0]^* d^3\mathbf{x}\\
&\quad+\int_{\Omega^{(s)}}\left[\omega^2\mathbf{g}(\mathbf{x},\mathbf{x}_1,\omega)\cdot\mathbf{g}^*(\mathbf{x},\mathbf{x}_2,\omega)\mathrm{Im}\,\rho\right.\\
&\quad\left.-\frac{\partial g_m(\mathbf{x},\mathbf{x}_1,\omega)}{\partial x_n}\frac{\partial g_j^*(\mathbf{x},\mathbf{x}_2,\omega)}{\partial x_k}\mathrm{Im}\,C_{jkmn}\right]d^3\mathbf{x}. \qquad (25)
\end{aligned}
$$

Identities of this kind, which relate a Green's function to an integral of a product of two Green's functions, are sometimes termed Ward identities.[60]

Now we introduce Green's functions for wave sources located within the solid. Let a source of external strain with $e_{mn}=\delta(\mathbf{x}-\mathbf{x}_1)\delta'(t)E_{mn}$ and $\mathbf{F}=0$ in the solid generate stresses and displacements with frequency spectra $\sigma_{jk}(\mathbf{x})=E_{mn}S_{mn,jk}(\mathbf{x},\mathbf{x}_1,\omega)$ and $\mathbf{w}(\mathbf{x})=E_{mn}\mathbf{T}_{mn}(\mathbf{x},\mathbf{x}_1,\omega)$ in the medium. Here E_{mn} is an arbitrary symmetric tensor. In particular, the frequency spectrum of the pressure generated by the external strain source at a point $\mathbf{x}$ in fluid is $p(\mathbf{x})=-S_{mn,jj}(\mathbf{x},\mathbf{x}_1,\omega)E_{mn}/3$. The tensor $S_{mn,jk}(\mathbf{x},\mathbf{x}_1,\omega)$ has the meaning of Green's tensor for stresses due to a CW source of external strain. It is invariant with respect to permutation of indices in the first or second pairs.

By applying the FRT (15) to wave fields generated by different sources of external strain in the original medium and in the medium with reversed flow, we obtain a reciprocity-type symmetry relation for the Green's tensor $S_{mn,jk}$ [compare with Eq. (23)]:

$$
S_{jk,mn}(\mathbf{x},\mathbf{x}_1,\omega)=\tilde{S}_{mn,jk}(\mathbf{x}_1,\mathbf{x},\omega). \qquad (26)
$$

Another important identity is obtained by applying the reciprocity relation (20) to wave fields generated in the original medium by two distinct sources of external strain with the frequency spectra $e_{mn}^{(r)}=-i\omega\delta(\mathbf{x}-\mathbf{x}_r)E_{mn}^{(r)}$, $\mathbf{F}^{(r)}=0$, $B^{(r)}=0$, and $r=1,2$. Again assuming, as in the derivation of Eq. (25), that $\Omega\equiv\Omega^{(f)}\cup\Omega^{(s)}$ represents the whole domain occupied by the fluid and the solid and that medium boundaries, if any, are free, rigid, or locally reacting surfaces with purely reactive impedance, from Eq. (20) we obtain

$$
\begin{aligned}
&-\frac{\omega}{2}\left[C_{jkmn}(\mathbf{x})\frac{\partial T_{ls,j}^*}{\partial x_k}(\mathbf{x},\mathbf{x}_2,\omega)\big|_{\mathbf{x}=\mathbf{x}_1}+C_{jkls}^*(\mathbf{x}_2)\frac{\partial T_{mn,j}}{\partial x_k}(\mathbf{x},\mathbf{x}_1,\omega)\big|_{\mathbf{x}=\mathbf{x}_2}\right]\\
&=\int_{\Omega^{(f)}}\left(\mathrm{Im}\frac{1}{\rho c^2}\right)\left[\frac{1}{3}S_{mn,jj}(\mathbf{x},\mathbf{x}_1,\omega)+\mathbf{T}_{mn}(\mathbf{x},\mathbf{x}_1,\omega)\cdot\nabla p_0\right]\left[\frac{1}{3}S_{ls,jj}(\mathbf{x},\mathbf{x}_2,\omega)+\mathbf{T}_{ls}(\mathbf{x},\mathbf{x}_2,\omega)\cdot\nabla p_0\right]^* d^3\mathbf{x}\\
&\quad+\int_{\Omega^{(s)}}\left[\omega^2\mathbf{T}_{mn}(\mathbf{x},\mathbf{x}_1,\omega)\cdot\mathbf{T}_{ls}^*(\mathbf{x},\mathbf{x}_2,\omega)\mathrm{Im}\,\rho-\frac{\partial T_{mn,j}(\mathbf{x},\mathbf{x}_1,\omega)}{\partial x_k}\frac{\partial T_{ls,j'}^*(\mathbf{x},\mathbf{x}_2,\omega)}{\partial x_{k'}}\mathrm{Im}\,C_{jkj'k'}(\mathbf{x})\right]d^3\mathbf{x}. \qquad (27)
\end{aligned}
$$

In the derivation of Eq. (27), we made use of arbitrariness of the symmetric tensors $E_{mn}^{(1)}$ and $E_{mn}^{(2)}$.

For a wave field generated by arbitrary distributed sources of external strain in the solid, volume velocity in the fluid, and external force, from the FRT (15) we find

$$
\begin{aligned}
\sigma_{mn}(\mathbf{x},\omega)=\frac{i}{\omega}\int d^3\mathbf{y}\left[\left(e_{jk}(\mathbf{y})-\frac{B(\mathbf{y})}{3}\delta_{jk}\right)\tilde{S}_{mn,jk}(\mathbf{y},\mathbf{x},\omega)\right.\\
\left.-\mathbf{F}(\mathbf{y})\cdot\tilde{\mathbf{T}}_{mn}(\mathbf{y},\mathbf{x},\omega)\right], \qquad (28)
\end{aligned}
$$

where the integral is taken over the entire volume occupied by the sources.

Let elastic waves be generated by random sources that are located in a fluid-solid domain Ω, have zero statistical mean, and are δ-correlated in space:

$$
\begin{aligned}
&\langle B(\mathbf{x})\rangle=0,\quad\langle\mathbf{F}(\mathbf{x})\rangle=0,\quad\langle e_{mn}(\mathbf{x})\rangle=0,\\
&\langle B(\mathbf{x})B^*(\mathbf{y})\rangle=Q^{(1)}(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}),\\
&\langle B(\mathbf{x})F_j^*(\mathbf{y})\rangle=Q_j^{(2)}(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}),\\
&\langle F_m(\mathbf{x})F_j^*(\mathbf{y})\rangle=Q_{mj}^{(3)}(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}),\\
&\langle e_{mn}(\mathbf{x})F_j^*(\mathbf{y})\rangle=Q_{mnj}^{(4)}(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}), \qquad (29)\\
&\langle e_{mn}(\mathbf{x})e_{jk}^*(\mathbf{y})\rangle=Q_{mnjk}^{(5)}(\mathbf{x})\delta(\mathbf{x}-\mathbf{y}).
\end{aligned}
$$

Here and below, angular brackets $\langle\cdot\rangle$ denote ensemble average. With B and e_{mn} being identically zero, respectively, within the solid and the fluid, the random source densities B

and e_{mn} are assumed to be uncorrelated. Due to the linearity of the problem, the oscillatory displacement, acoustic pressure in the fluid, and stresses in the solid in the resulting random wave field are zero on average. For a two-point correlation function of acoustic pressure in a moving fluid, from Eqs. (24) and (29) we find

$$\langle p(\mathbf{x}_1)p^*(\mathbf{x}_2)\rangle = \omega^{-2}\int_\Omega d^3\mathbf{y}[Q^{(1)}(\mathbf{y})\tilde{G}(\mathbf{y},\mathbf{x}_1)\tilde{G}^*(\mathbf{y},\mathbf{x}_2) - Q_j^{(2)}(\mathbf{y})\tilde{G}(\mathbf{y},\mathbf{x}_1)\tilde{g}_j^*(\mathbf{y},\mathbf{x}_2) - Q_j^{(2)*}(\mathbf{y})\tilde{G}^*(\mathbf{y},\mathbf{x}_2)\tilde{g}_j(\mathbf{y},\mathbf{x}_1)$$
$$+ Q_{mj}^{(3)}(\mathbf{y})\tilde{g}_m(\mathbf{y},\mathbf{x}_1)\tilde{g}_j^*(\mathbf{y},\mathbf{x}_2) - Q_{mnl}^{(4)}(\mathbf{y})\tilde{s}_{mn}(\mathbf{y},\mathbf{x}_1)\tilde{g}_l^*(\mathbf{y},\mathbf{x}_2) - Q_{mnl}^{(4)*}(\mathbf{y})\tilde{s}_{mn}^*(\mathbf{y},\mathbf{x}_1)\tilde{g}_l(\mathbf{y},\mathbf{x}_2)$$
$$+ Q_{mnjk}^{(5)}(\mathbf{y})\tilde{s}_{mn}(\mathbf{y},\mathbf{x}_1)\tilde{s}_{jk}^*(\mathbf{y},\mathbf{x}_2)]. \tag{30}$$

Here and below, where it cannot lead to confusion, the argument ω is suppressed for brevity. Similar expressions can be obtained for correlation of other observables in the fluid and the solid. In particular, when external forces are uncorrelated with sources of the external strain in the solid (and, hence, $Q_{mnj}^{(4)} \equiv 0$) for the cross-correlation of stresses within the solid, from Eqs. (28) and (29) we find

$$\langle \sigma_{mn}(\mathbf{x}_1)\sigma_{jk}^*(\mathbf{x}_2)\rangle = \omega^{-2}\int_\Omega \{[Q_{lqrs}^{(5)}(\mathbf{y})$$
$$+ Q^{(1)}\delta_{lq}\delta_{rs}/9]\tilde{S}_{mn,lq}(\mathbf{y},\mathbf{x}_1)\tilde{S}_{jk,rs}^*(\mathbf{y},\mathbf{x}_2)$$
$$- Q_s^{(2)}(\mathbf{y})\tilde{S}_{mn,qq}(\mathbf{y},\mathbf{x}_1)\tilde{T}_{jk,s}^*(\mathbf{y},\mathbf{x}_2)/3$$
$$- Q_s^{(2)*}(\mathbf{y})\tilde{S}_{jk,qq}^*(\mathbf{y},\mathbf{x}_2)\tilde{T}_{mn,s}(\mathbf{y},\mathbf{x}_1)/3$$
$$+ Q_{ls}^{(3)}(\mathbf{y})\tilde{T}_{mn,l}(\mathbf{y},\mathbf{x}_1)\tilde{T}_{jk,s}^*(\mathbf{y},\mathbf{x}_2)\}d^3\mathbf{y}. \tag{31}$$

Note that the Green's functions and tensors in the integrands in Eqs. (24) and (28) [and, consequently, in Eqs. (30) and (31)] correspond to wave fields in the medium with reversed flow. Integration in Eqs. (24), (28), (30), and (31) is over the receiver, rather than the source, position. The latter property of Eqs. (30) and (31) greatly simplifies derivation of local relationships between noise cross-correlation and deterministic Green's functions and tensors, which is outlined below in Sec. VI.

V. THERMAL NOISE AND THE FLUCTUATION-DISSIPATION THEOREM

A thermal acoustic field in fluids and solids is produced by the chaotic thermal motion of microscopic particles such as atoms, ions, and electrons. The FDT allows one to represent acoustic noise at a state close to thermal equilibrium as a field due to effective, random, macroscopic wave sources. Let small perturbations of a continuous system be described by a set $\{\xi_\beta(\mathbf{x},t)\}, \beta=1,2,3,\ldots,N$ of generalized coordinates, and macroscopic equations of motion can be written in terms of spectral components of the generalized coordinates as follows:

$$\sum_{\gamma=1}^{N}\hat{K}_{\beta\gamma}(\mathbf{x},\omega)\xi_\gamma(\mathbf{x},\omega) = f_\beta(\mathbf{x},\omega), \quad \beta=1,2,\ldots,N. \tag{32}$$

Here $\hat{K}_{\beta\gamma}(\mathbf{x},\omega)$ is a linear operator, and $f_\beta(\mathbf{x},\omega)$ are spectral components of generalized forces. The generalized forces and generalized coordinates are also related by the requirement that the statistical average of volume density of the power created by a force $f_\beta(\mathbf{x},t)$ is $\langle f_\beta(\mathbf{x},t)\,\partial\xi_\beta(\mathbf{x},t)/\partial t\rangle$. According to the FDT, thermal fluctuations of the system can be viewed as generated by random generalized forces with zero mean and the following correlation functions (Ref. 2, p. 119; Ref. 3, p. 273; Ref. 55, p. 389):

$$\langle f_\beta(\mathbf{x})f_\gamma^*(\mathbf{y})\rangle = -\frac{i\Theta}{2\pi\omega}[\hat{K}_{\gamma\beta}^*(\mathbf{x},\omega) - \hat{K}_{\beta\gamma}(\mathbf{y},\omega)]\delta(\mathbf{x}-\mathbf{y}),$$
$$\Theta = \frac{\hbar\omega}{2}\coth\frac{\hbar\omega}{2\kappa T}, \tag{33}$$

where T is temperature, Θ is the mean energy of a quantum oscillator, and $\hbar$ and κ are Planck's and Boltzmann's constants. In the classical (or high-temperature) limit, where $\hbar\omega \ll \kappa T$, $\Theta(T,\omega) \approx \kappa T$.

To characterize the thermal noise sources located within a moving dissipative fluid, we choose[43] $\xi_j = w_j$, $j=1,2,3$; $\xi_4 = p$. From the expression[57] $\int d^3\mathbf{x}[\mathbf{F}(\mathbf{x},t)\cdot\partial\mathbf{w}(\mathbf{x},t)/\partial t + p(\mathbf{x},t)\,\partial B(\mathbf{x},t)/\partial t]$ for acoustic energy production by sound sources, it follows that the generalized forces corresponding to the chosen generalized coordinates are $f_j = F_j$, $j=1,2,3$; $f_4 = -B$. Statistical properties of these sources[43] follow Eq. (29) with

$$Q^{(1)} = \frac{\Theta}{\pi\omega}\mathrm{Im}\frac{1}{\rho c^2}, \quad Q_m^{(2)} = -\frac{\Theta}{\pi\omega}\frac{\partial p_0}{\partial x_m}\mathrm{Im}\frac{1}{\rho c^2}, \tag{34}$$
$$Q_{mn}^{(3)} = \frac{\Theta}{\pi\omega}\frac{\partial p_0}{\partial x_m}\frac{\partial p_0}{\partial x_n}\mathrm{Im}\frac{1}{\rho c^2}, \quad \mathbf{x}\in\Omega^{(f)},$$

when wave dissipation in the fluid is described by its complex compressibility. Note that in a generic inhomogeneous flow, the sources of extraneous force and volume velocity prove to be correlated. This is in contrast to the *ad hoc* source model assumed in Ref. 42.

To characterize the thermal noise sources located in a solid, we choose $\xi_j = w_j$, $j=1,2,3$; $\xi_4 = -\sigma_{11}$, $\xi_5 = -(\sigma_{12}+\sigma_{21})$, $\xi_6 = -(\sigma_{13}+\sigma_{31})$, $\xi_7 = -\sigma_{22}$, $\xi_8 = -(\sigma_{23}+\sigma_{32})$, and $\xi_9 = -\sigma_{33}$ as generalized coordinates. Following Ref. 58, it is easy to show that the rate of wave energy production by sources of external force and strain in a solid is given by the expression $\int d^3\mathbf{x}[\mathbf{F}(\mathbf{x},t)\cdot\partial\mathbf{w}(\mathbf{x},t)/\partial t + e_{mn}(\mathbf{x},t)\,\partial\sigma_{mn}(\mathbf{x},t)/\partial t]$. Hence, the generalized forces corresponding to the chosen generalized coordinates are $f_j = F_j$, $j=1,2,3$; $f_4 = e_{11}$, $f_5 = e_{12}+e_{21}$, $f_6 = e_{13}+e_{31}$, $f_7 = e_{22}$, $f_8 = e_{23}+e_{32}$, and $f_9 = e_{33}$. The operators $\hat{K}_{\beta\gamma}$, $\beta,\gamma = 1,2,\ldots,9$ introduced in Eq. (32) are found from Eqs. (6) and (8): $\hat{K}_{jk} = -\rho\omega^2\delta_{jk}$, $\hat{K}_{j4} = \delta_{j1}\partial/\partial x_1$, $\hat{K}_{j5} = \delta_{j1}\partial/\partial x_2 + \delta_{j2}\partial/\partial x_1$, $j,k = 1,2,3,\ldots,\hat{K}_{56} = -4\alpha_{1213},\ldots,\hat{K}_{98} = -2\alpha_{2333},\hat{K}_{99} = -\alpha_{3333}$. Using Eq. (33) to calculate correlation functions of the volume densities of the effective sources $\mathbf{F}$ and e_{mn} of thermal fluctuations in the solid, we arrive at Eq. (29) with

$$Q^{(3)}_{mn} = \frac{\Theta\omega}{\pi}\delta_{mn}\,\mathrm{Im}\,\rho, \quad Q^{(4)}_{mnj} = 0, \tag{35}$$

$$Q^{(5)}_{mnjk} = \frac{\Theta}{\pi\omega}\mathrm{Im}\,\alpha_{mnjk}, \quad \mathbf{x}\in\Omega^{(s)}.$$

An inspection shows that Eq. (35) reduces to known results in the two particular cases previously considered in the literature, namely, for waves in a motionless fluid,[43] where $\alpha_{mmnn} = (\rho c^2)^{-1}$, and for waves in a homogeneous isotropic solid with $\mathrm{Im}\,\rho = 0$.[62] Note that even in the fluid case, where wave dissipation is described by two scalar quantities $\mathrm{Im}\,\rho$ and $\mathrm{Im}(\rho c^2)^{-1}$, the volume density of the noise sources is generally not proportional to the local value of the plane-wave attenuation coefficient, unlike in the simple phenomenological model of Ref. 39.

Although Eq. (33) strictly applies only at thermal equilibrium, resulting expressions for effective sources of thermal noise hold also when the temperature T varies as long as temperature gradients are sufficiently small so that the contribution of non-equilibrium processes remains negligible (Ref. 3, p. 128). Then Eqs. (34) and (35) allow one to calculate two-point statistical moments of thermal wave fields in the fluid and in the solid in rather general systems, for instance, in the inhomogeneous flow of a cold gas past heated solid bodies. Similar calculations for a number of simple geometries, e.g., waves due to delta-correlated random sources in a sphere or in a half-space with a plane boundary, have been carried out[2,3] in studies of thermal electromagnetic fields. The necessary calculations and especially their results simplify greatly when temperature is constant throughout the domain occupied by the effective sources of thermal noise. This important special case is considered in the next section.

VI. GREEN'S FUNCTION RETRIEVAL FROM NOISE CROSS-CORRELATION

Consider the two-point correlation of the spectral densities of pressure fluctuations within the fluid in a fluid-solid system in thermal equilibrium. We have shown in Sec. V that the thermal fluctuations can be viewed as generated by delta-correlated random sources of volume velocity and external force located within the fluid and by sources of external force and strain located within the solid. Taking into account that the mean energy Θ [Eq. (33)] of a quantum oscillator does not depend on position, from Eqs. (30), (34), and (35) we find

$$\begin{aligned}\langle p(\mathbf{x}_1,\omega)p^*(\mathbf{x}_2,\omega)\rangle &= \frac{\Theta}{\pi\omega^3}\left\{\int_{\Omega^{(f)}}\left(\mathrm{Im}\frac{1}{\rho c^2}\right)[G(\mathbf{y},\mathbf{x}_1,\omega)+\mathbf{g}(\mathbf{y},\mathbf{x}_1,\omega)\cdot\nabla p_0]\right.\\ &\times[G(\mathbf{y},\mathbf{x}_2,\omega)+\mathbf{g}(\mathbf{y},\mathbf{x}_2,\omega)\cdot\nabla p_0]^* d^3\mathbf{y}\\ &+\int_{\Omega^{(s)}}[\omega^2\mathbf{g}(\mathbf{y},\mathbf{x}_1,\omega)\cdot\mathbf{g}^*(\mathbf{y},\mathbf{x}_2,\omega)\mathrm{Im}\,\rho\\ &\left.+\tilde{s}_{mn}(\mathbf{y},\mathbf{x}_1,\omega)\tilde{s}^*_{jk}(\mathbf{y},\mathbf{x}_2,\omega)\mathrm{Im}\,\alpha_{jkmn}]d^3\mathbf{y}\right\}.\end{aligned} \tag{36}$$

This result extends earlier calculations of the thermal noise cross-correlation in homogeneous or weakly inhomogeneous quiescent fluid[50,62] and in an arbitrary inhomogeneous moving fluid[43] to flow-structure problems. As long as thermal fluctuations have Gaussian statistics, knowledge of the correlation function (36) provides a full description of the random field $p(\mathbf{x},\omega)$ and enables one to calculate any statistical moment of the pressure fluctuations.[3]

From the property (10) of the elastic modulus and compliance tensors and Eq. (7), which relate displacements $\mathbf{g}$ and stresses s_{mn} due to the same source within the fluid, it follows that $s_{mn}(\mathbf{y},\mathbf{x}_1)s^*_{jk}(\mathbf{y},\mathbf{x}_2)\mathrm{Im}\,\alpha_{jkmn} = -(\partial g_m/\partial y_n)(\mathbf{y},\mathbf{x}_1)\times(\partial g^*_j/\partial y_k)(\mathbf{y},\mathbf{x}_2)\mathrm{Im}\,C_{jkmn}$. Then, using the integral identity (25) and the symmetry property (23) of the Green's functions, from Eq. (36) we obtain

$$\begin{aligned}\langle p(\mathbf{x}_1,\omega)p^*(\mathbf{x}_2,\omega)\rangle = &-(2\pi\omega^2)^{-1}\Theta[G(\mathbf{x}_1,\mathbf{x}_2,\omega)\\ &+G^*(\mathbf{x}_2,\mathbf{x}_1,\omega)].\end{aligned} \tag{37}$$

Thus, in the frequency domain, the cross-correlation of thermal noise equals the product of a temperature-dependent function, which does not depend on coordinates, and a sum of the CW Green's functions, which describe sound propagation in opposite directions between the two points. In the particular case of motionless media, the Green's function G is reciprocal according to Eq. (23), and the sum reduces to twice the real part of the Green's function. The simple and general relation (37) between the two-point correlation of thermal pressure fluctuations and the deterministic Green's function, i.e., the pressure field due to a point source of volume velocity, in fluid-solid systems with flow does not differ from the result[43] derived previously in the fluid case.

The two-point correlation of the spectral densities of stress fluctuations within the solid in a fluid-solid system in thermal equilibrium can be calculated using Eqs. (31), (34), and (35). With Θ [Eq. (33)] being independent of position,

we find

$$\langle \sigma_{mn}(\mathbf{x}_1,\omega)\sigma^*_{jk}(\mathbf{x}_2,\omega)\rangle = \frac{\Theta}{\pi\omega^3}\left\{\int_{\Omega^{(f)}} \mathrm{Im}\frac{1}{\rho c^2}\left[\frac{1}{3}\tilde{S}_{mn,ll}(\mathbf{y},\mathbf{x}_1,\omega) + \tilde{\mathbf{T}}_{mn}(\mathbf{y},\mathbf{x}_1,\omega)\cdot\nabla p_0\right]\times\left[\frac{1}{3}\tilde{S}_{jk,rr}(\mathbf{y},\mathbf{x}_2,\omega) + \tilde{\mathbf{T}}_{jk}(\mathbf{y},\mathbf{x}_2,\omega)\cdot\nabla p_0\right]^* d^3\mathbf{y} + \int_{\Omega^{(s)}} [\omega^2\tilde{\mathbf{T}}_{mn}(\mathbf{y},\mathbf{x}_1,\omega)\cdot\tilde{\mathbf{T}}^*_{jk}(\mathbf{y},\mathbf{x}_2,\omega)\mathrm{Im}\ \rho + \tilde{S}_{mn,lq}(\mathbf{y},\mathbf{x}_1,\omega)\tilde{S}^*_{jk,rs}(\mathbf{y},\mathbf{x}_2,\omega)\mathrm{Im}\ \alpha_{lqrs}(\mathbf{y})]d^3\mathbf{y}\right\}. \quad (38)$$

In order to simplify this result, we multiply Eq. (38) by arbitrary symmetric tensors E_{mn} and E^*_{jk}, use Eq. (7) to express $\tilde{S}_{mn,lq}$ and $\tilde{S}^*_{jk,rs}$ in the integral over $\Omega^{(s)}$ in terms of $\tilde{\mathbf{T}}_{mn}$ and $\tilde{\mathbf{T}}^*_{jk}$:

$$E_{mn}\tilde{S}_{mn,lq}(\mathbf{y},\mathbf{x}_1,\omega) = C_{lqrs}(\mathbf{y})[E_{mn}\,\partial T_{mn,m'}(\mathbf{y},\mathbf{x}_1,\omega)/\partial y_{n'} - i\omega E_{m'n'}\delta(\mathbf{y}-\mathbf{x}_1)],$$

$$E^*_{jk}\tilde{S}^*_{jk,rs}(\mathbf{y},\mathbf{x}_2,\omega) = C_{rsj'k'}(\mathbf{y})[E^*_{jk}\,\partial T^*_{jk,j'}(\mathbf{y},\mathbf{x}_2,\omega)/\partial y_{k'} - i\omega E_{j'k'}\delta(\mathbf{y}-\mathbf{x}_2)],$$

and apply Eq. (10) to relate imaginary parts of the material tensors $\alpha_{lqr\,s}$ and C_{lqrs}. After some algebra, involving the symmetry property (26) and the integral identity (27) for the Green's tensors $S_{mn,lq}$ and $\mathbf{T}_{mn}$, from Eq. (38) we obtain the simple relation

$$\langle \sigma_{mn}(\mathbf{x}_1,\omega)\sigma^*_{jk}(\mathbf{x}_2,\omega)\rangle = -(2\pi\omega^2)^{-1}\Theta[S_{jk,mn}(\mathbf{x}_1,\mathbf{x}_2,\omega) + S^*_{mn,jk}(\mathbf{x}_2,\mathbf{x}_1,\omega)] \quad (39)$$

between the cross-correlation function of thermal fluctuations of stresses within the solid and the deterministic Green's tensor corresponding to sources of the external strain. Equation (39) extends to flow-structure problems the Rytov's calculation[62] of the thermal noise cross-correlation in an isotropic homogeneous solid. Note that as in Eq. (37), the right-hand side of Eq. (39) contains Green's tensors describing wave propagation in opposite directions between the points $\mathbf{x}_1$ and $\mathbf{x}_2$.

Remarkably, the parameters specifying dissipative properties of the medium drop out from the final results, Eqs. (37) and (39). These equations express the elastic wave counterpart of the result established much earlier for equilibrium thermal electromagnetic fields (Ref. 2, Sec. 3.9; Ref. 3, Sec. 15) in motionless dielectrics.

In the high-temperature limit, Θ becomes independent of frequency: $\Theta(T,\omega)\approx\kappa T$, and, as pointed out in Ref. 43, Eq. (37) can be concisely written in the time domain. If one defines the cross-correlation function $C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,\tau)$ of statistically stationary pressure fluctuations within the fluid and the power spectrum $C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,\omega)$ of the fluctuations in a standard way as

$$C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,\tau) \equiv \langle p(\mathbf{x}_1,t-\tau)p(\mathbf{x}_2,t)\rangle = \int_{-\infty}^{+\infty} C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,\omega)e^{-i\omega\tau}d\omega, \quad (40)$$

then $C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,\omega)=\langle p^*(\mathbf{x}_1,\omega)p(x_2,\omega)\rangle$, and from Eq. (37) we find

$$(\partial^2/\partial\tau^2)C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,\tau) = (2\pi)^{-1}\Theta[G(\mathbf{x}_1,\mathbf{x}_2,-\tau) + G(\mathbf{x}_2,\mathbf{x}_1,\tau)]. \quad (41)$$

Because of causality, $G(\mathbf{x},\mathbf{y},t)=0$ at $t<0$. Hence,

$$G(\mathbf{x}_2,\mathbf{x}_1,\tau) = \frac{2\pi}{\kappa T}\frac{\partial^2}{\partial\tau^2}C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,\tau), \quad (42)$$

$$G(\mathbf{x}_1,\mathbf{x}_2,\tau) = \frac{2\pi}{\kappa T}\frac{\partial^2}{\partial\tau^2}C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,-\tau), \quad \tau>0.$$

In motionless media, cross-correlation function $C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,\tau)$ is an even function of time delay τ; in moving fluid, according to Eq. (42), the difference $C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,\tau) - C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,-\tau)$ provides a measure of acoustic nonreciprocity.

Quite similarly, the Green's tensors $S_{mn,jk}(\mathbf{x}_1,\mathbf{x}_2,t)$ in the time domain can be retrieved from measurements of the cross-correlation function

$$C^{(\sigma)}_{mnjk}(\mathbf{x}_1,\mathbf{x}_2,\tau) \equiv \langle \sigma_{mn}(\mathbf{x}_1,t-\tau)\sigma_{jk}(\mathbf{x}_2,t)\rangle = \int_{-\infty}^{+\infty} C^{(\sigma)}_{mnjk}(\mathbf{x}_1,\mathbf{x}_2,\omega)e^{-i\omega\tau}d\omega \quad (43)$$

of fluctuations of the stress tensor components within the solid. Because of causality, $S_{mn,jk}(\mathbf{x}_1,\mathbf{x}_2,t)=0$ at $t<0$. Then, from Eqs. (39) and (43) it follows that

$$S_{jk,mn}(\mathbf{x}_2,\mathbf{x}_1,\tau) = \frac{2\pi}{\kappa T}\frac{\partial^2}{\partial\tau^2}C^{(\sigma)}_{mn,jk}(\mathbf{x}_1,\mathbf{x}_2,\tau),$$

$$S_{jk,mn}(\mathbf{x}_1,\mathbf{x}_2,\tau) = \frac{2\pi}{\kappa T}\frac{\partial^2}{\partial\tau^2}C^{(\sigma)}_{mn,jk}(\mathbf{x}_1,\mathbf{x}_2,-\tau), \quad \tau>0. \quad (44)$$

Note that in a solid in a mechanical contact with a fluid flow, generally $C^{(\sigma)}_{mn,jk}(\mathbf{x}_1,\mathbf{x}_2,\tau)\neq C^{(\sigma)}_{mn,jk}(\mathbf{x}_1,\mathbf{x}_2,-\tau)$ and elastic wave propagation is nonreciprocal, although the solid itself is macroscopically motionless. Like Eq. (38) used in the derivation of Eq. (44), the latter applies also in the case where one of the points $\mathbf{x}_1$ and $\mathbf{x}_2$ is located within the fluid while the other is within the solid.

The results (42) and (44) will hold at an arbitrary temperature T if one replaces the noise correlation functions by their appropriately filtered versions, the filter shape being uniquely determined by the frequency dependence $\Theta(T,\omega)$ [Eq. (33)] of the mean energy of a quantum oscillator.

VII. DISCUSSION

In view of the extensive earlier work reviewed in Sec. I, it is now widely expected that a two-point correlation function of ambient noise provides an estimate of or is "essentially equal"[60] to a deterministic Green's function. The actual questions, then, are: under what conditions does the relation between the cross-correlation and Green's functions take place, whether it is exact or approximate, and what is the nature of the approximation. Answers to these questions are system-specific. In Secs. II–VI, we considered thermal fluctuations in rather general fluid-solid systems with stationary inhomogeneous flow. Such mechanical systems, while encountered in a wide range of applications, apparently do not belong to the class of systems covered by recent "unified" derivations[42,60,63] of the relation between the noise cross-correlation and Green's function. Moreover, our approach differs from some earlier work[6,39,42,64,65] in that, instead of *assuming* noise source properties, which simplify derivations and/or lead to a compact relation between noise cross-correlations and deterministic Green's functions, we *derive* effective macroscopic sources of thermal noise from the first principles.

In the analysis of thermal fluctuations in Secs. V and VI, we have purposefully avoided energy equipartitioning arguments.[8,9,41,44,63] Instead, the FDT is used in the spectral form proposed by Rytov[1–3,56] to determine local densities of equivalent macroscopic sources of thermal noise. This leads to Eqs. (36) and (38), which describe statistics of thermal fluctuations in heterogeneous systems with *variable* temperature. In the special case of position-independent temperature, integral expressions (36) and (38) for the thermal noise cross-correlation simplify greatly and allow for exact local relations (37), (42), and (44) with deterministic Green's functions (tensors).

Weaver and Lobkis[7,8] presented an elegant proof of equivalence between pulse echo response, defined as voltage in a receiver circuit of an acoustic transducer after emission of a short pulse by the same transducer, and cross-correlation of thermal noise in the receiver circuit of the transducer. The derivation assumes thermal equilibrium, relies on the Callen–Welton formulation[66] of the FDT, and is valid in the limit of vanishing resistive part of electro-mechanical impedance of the transducer. The derivation[7,8] is remarkably simple and does not require an actual analysis of fields of elastic waves. A difficulty of extending the derivation to noise cross-correlation was pointed out.[7,8] In contrast to this approach, the results obtained in Secs. V and VI are formulated in terms of acoustic observables and apply to both cross-correlation and autocorrelation functions of noise.

This paper extends earlier analyses of thermal noise in inhomogeneous moving fluids[43] and motionless fluids[50] and solids[7,8,62] to bounded or unbounded coupled fluid–anisotropic solid systems with arbitrary inhomogeneous fluid flow; the fluid and the solid can be arbitrarily inhomogeneous, and neither fluid nor solid domains are required to be simply connected or finite. Remarkably, at thermal equilibrium (but not in the general case of spatially varying temperature), the relation between the cross-correlation of thermal pressure fluctuations and the deterministic Green's function is unaffected by the presence of the solid. Unlike an earlier theory[7,8] of thermal noise correlation in motionless media, our results apply equally to open systems, waveguides, and resonators.

It should be emphasized that the appearance of the relation between the noise correlation function and the Green's function depends on the definition of the latter and on the choice of a physical observable used to represent the noise. For instance, if the Green's function were defined as the pressure field due to a source of volume velocity with $B=\delta(\mathbf{x}-\mathbf{x}_1)\delta(t)$ instead of $B=\delta(\mathbf{x}-\mathbf{x}_1)\delta'(t)$, first-order derivatives with respect to τ would appear in Eq. (42) instead of the second-order derivatives. If the Green's function were defined as the acoustic pressure due to a source with $B=\delta(\mathbf{x}-\mathbf{x}_1)H(t)$, where $H(t)$ denotes the step function, the Green's function $G(\mathbf{x}_2,\mathbf{x}_1,t)$ would be proportional to the correlation function $C^{(p)}(\mathbf{x}_1,\mathbf{x}_2,t)$ of thermal pressure fluctuations at high temperatures. Furthermore, if the Green's function were defined as the acoustic pressure within the fluid due to a source of an external force, rather than a source of volume velocity, no relation as simple as Eq. (42) would hold in the general case of inhomogeneous fluid flow with or without solid inclusions. For the simple relations (44) between cross-correlation of noise within the solid and the deterministic Green's tensor $S_{jk,mn}$ to be valid, the noise has to be characterized in terms of the stress tensor fluctuations.

The practical significance of our theoretical findings lies in their possible application to passive acoustic characterization of fluid-solid systems, including monitoring of the physical parameters of inhomogeneous flows. In many geophysical, engineering, and biomedical applications, the flow velocity is small compared to the phase speed of elastic waves. For instance, the ratio of velocity of ocean currents and winds in the troposphere to sound speed typically does not exceed 3×10^{-4} and 5×10^{-2}, respectively. Admissible absolute errors in flow velocity measurements are often much smaller than uncertainty in knowledge of the sound speed. The difficulties of remote sensing of slow flows in uncertain environments are overcome by measurements of acoustic nonreciprocity in reciprocal transmission experiments, where time-synchronized acoustic transceivers, i.e., a combination of a sound source and receiver(s), are employed to generate and receive signals propagating in opposite directions.[67,68] Non-reciprocity of acoustic observables, such as the difference in travel times or phases of waves propagating between two points in opposite directions, is induced by fluid flow, vanishes in quiescent media, and is insensitive to uncertainties in sound speed and system geometry. Reciprocal transmission experiments enable active acoustic tomography of slow flows.[67,68] Equations (37), (39), (42), and (44) demonstrate that all the information normally obtained via active reciprocal transmissions is retrievable from cross-correlation of thermal noise. Note that one can make noise measurements in a motionless part of the fluid or within the solid (e.g., within or on the external wall of a pipe carrying a fluid flow). Thus, ambient noise cross-correlation can be employed for passive acoustic characterization of inhomogeneous flows in fluid-solid environments.

Passive acoustic techniques offer important advantages over remote sensing of ocean and atmosphere with active techniques: low cost (a receiver substitutes a technologically much more complicated transceiver) and possibility of noninvasive (in particular, avoiding any harm to marine life potentially associated with powerful underwater sound sources and likely annoyance to local populations associated with active acoustic tomography of the atmosphere) and clandestine measurements in denied areas. Furthermore, the passive systems allow for exploitation of extremely broad bandwidth of ambient noise, which exceeds by far the bandwidth of available non-explosive manmade sound sources used in remote sensing; much longer periods of autonomous operation due to drastic reduction in power consumption; and increased spatial resolution of measurements due to the greater number of paths along which the Green's function is evaluated, compared to an active system with the same number of sensors. The chief potential benefit for biomedical applications lies in the noninvasive nature of passive *in vivo* measurements.[35,49,53] The key practical limitation of the approach is the requirement of a sufficiently long averaging time necessary to evaluate the correlation function of noise, during which the system needs to remain time-invariant.

The theoretical results presented above can be extended to ambient noise of non-thermal origin, which is due to random sources distributed on boundaries and interfaces or within the solid and the fluid. When the density of the random sources is not necessarily connected to the local dissipative properties of the medium, no exact local relation between the noise cross-correlation and deterministic Green's functions generally holds. However, an asymptotic estimate of the Green's functions can be obtained from the noise cross-correlation of high-frequency waves when the wavelength is small compared to spatial scales of variability of the parameters of the medium and statistical moments of densities of the random sources. This extension of the theory will be presented in a separate paper.

ACKNOWLEDGMENTS

This work was supported in part by the Office of Naval Research through Grant No. N00014-08-1-0100. Discussions with M. Charnotskii, V. G. Irisov, O. I. Lobkis, A. G. Voronovich, and R. I. Weaver are gratefully acknowledged.

[1]S. M. Rytov, "On thermal agitation in distributed systems," Sov. Phys. Dokl. **1**, 555–559 (1956).
[2]M. L. Levin and S. M. Rytov, *A Theory of Equilibrium Thermal Fluctuations in Electrodynamics* (Nauka, Moscow, 1967) (in Russian).
[3]S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radiophysics. 3: Elements of Random Fields* (Springer, New York, 1989), Chaps. 1 and 3.
[4]J. F. Claerbout, "Synthesis of a layered medium from its acoustic transmission response," Geophysics **33**, 264–269 (1968).
[5]J. Rickett and J. Claerbout, "Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring," The Leading Edge **18**, 957–960 (1999).
[6]K. Wapenaar, "Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation," Phys. Rev. Lett. **93**, 254301 (2004).
[7]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[8]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).
[9]O. I. Lobkis and R. L. Weaver, "On the emergence of the Green's function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**, 3011–3017 (2001).
[10]M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[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]N. M. Shapiro, M. Campillo, L. Stehly, and M. Ritzwoller, "High resolution surface wave tomography from ambient seismic noise," Science **307**, 1615–1618 (2005).
[13]K. G. Sabra, P. Gerstoft, and P. Roux, "Extracting time-domain Green's function from ambient seismic noise," Geophys. Res. Lett. **32**, L03310 (2005).
[14]R. L. Weaver, "Information from seismic noise," Science **307**, 1568–1569 (2005).
[15]M. Campillo, "Phase and correlation in 'random' seismic fields and the reconstruction of the Green function," Pure Appl. Geophys. **163**, 475–502 (2006).
[16]P. Gouédard, L. Stehly, F. Brenguier, M. Campillo, Y. Colin de Verdière, E. Larose, L. Margerin, P. Roux, F. J. Sánchez-Sesma, N. M. Shapiro, and R. L. Weaver, "Cross-correlation of random fields: Mathematical approach and applications," Geophys. Prospect. **56**, 375–393 (2008).
[17]K. G. Sabra, P. Roux, P. Gerstoft, W. A. Kuperman, and M. C. Fehler, "Extracting coherent coda arrivals from cross-correlations of long period seismic waves during the Mount St. Helens 2004 eruption," Geophys. Res. Lett. **33**, L06313 (2006).
[18]Y. Yang, M. H. Ritzwoller, A. L. Levshin, and N. M. Shapiro, "Ambient noise Rayleigh wave tomography across Europe," Geophys. J. Int. **168**, 259–274 (2007).
[19]P. Gouédard, C. Cornou, and P. Roux, "Phase-velocity dispersion curves and small-scale geophysics using noise correlation slantstack technique," Geophys. J. Int. **172**, 971–981 (2008).
[20]P. Gouédard, P. Roux, and M. Campillo, "Small-scale seismic inversion using surface waves extracted from noise cross correlation," J. Acoust. Soc. Am. **123**, EL26–EL31 (2008).
[21]T. L. Duvall, Jr., S. M. Jefferies, J. W. Harvey, and M. A. Pomerantz, "Time-distance helioseismology," Nature (London) **362**, 430–432 (1993).
[22]P. M. Giles, T. L. Duvall, Jr., P. H. Scherrer, and R. S. Bogart, "A subsurface flow of material from the Sun's equator to its poles," Nature (London) **390**, 52–54 (1997).
[23]J. E. Rickett and J. F. Claerbout, "Calculation of the Sun's impulse response by multi-dimensional spectral factorization," Sol. Phys. **192**, 203–210 (2000).
[24]P. Roux, W. A. Kuperman, and the NPAL Group, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[25]K. G. Sabra, P. Roux, A. M. Thode, G. L. D'Spain, W. S. Hodgkiss, and W. A. Kuperman, "Using ocean ambient noise for array self-localization and self-synchronization," IEEE J. Ocean. Eng. **30**, 338–347 (2005).
[26]K. G. Sabra, P. Roux, and W. A. Kuperman, "Arrival-time structure of the time-averaged ambient noise cross-correlation function in an oceanic waveguide," J. Acoust. Soc. Am. **117**, 164–174 (2005).
[27]L. A. Brooks and P. Gerstoft, "Ocean acoustic interferometry," J. Acoust. Soc. Am. **121**, 3377–3385 (2007).
[28]P. Gerstoft, W. S. Hodgkiss, M. Siderius, and C. H. Harrison, "Passive fathometer processing," J. Acoust. Soc. Am. **123**, 1297–1305 (2008).
[29]C. H. Harrison, "Target detection and location with ambient noise," J. Acoust. Soc. Am. **123**, 1834–1837 (2008).
[30]C. Farrar and G. James, "System identification from ambient vibration measurements on a bridge," J. Sound Vib. **205**, 1–18 (1997).
[31]R. Snieder and E. Cafak, "Extracting the building response using seismic interferometry; theory and application to the Millikan Library in Pasadena, California," Bull. Seismol. Soc. Am. **96**, 586–598 (2006).
[32]K. G. Sabra, E. S. Winkel, D. A. Bourgoyne, B. R. Elbing, S. L. Ceccio, M. Perlin, and D. R. Dowling, "On using cross-correlation of turbulent flow-induced ambient vibrations to estimate the structural impulse response. Application to structural health monitoring," J. Acoust. Soc. Am. **121**, 1987–2005 (2007).
[33]K. G. Sabra, A. Srivastava, F. L. di Scalea, I. Bartoli, P. Rizzo, and S. Conti, "Structural health monitoring by extraction of coherent guided

waves from diffuse fields," J. Acoust. Soc. Am. **123**, EL8–EL13 (2008).
[34]V. I. Mirgorodskii, V. V. Gerasimov, and S. V. Peshin, "Experimental studies of passive correlation tomography of incoherent acoustic sources in the Megahertz frequency band," Acoust. Phys. **52**, 606–612 (2006).
[35]K. G. Sabra, S. Conti, P. Roux, and W. A. Kuperman, "Passive *in vivo* elastography from skeletal muscle noise," Appl. Phys. Lett. **90**, 194101 (2007).
[36]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).
[37]A. Derode, E. Larose, M. Tanter, J. de Rosny, A. Tourin, M. Campillo, and M. Fink, "Recovering the Green's function from field-field correlations in an open scattering medium," J. Acoust. Soc. Am. **113**, 2973–2976 (2003).
[38]R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).
[39]R. Snieder, "Extracting the Green's function of attenuating heterogeneous acoustic media from uncorrelated waves," J. Acoust. Soc. Am. **121**, 2637–2643 (2007).
[40]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).
[41]R. L. Weaver and O. I. Lobkis, "Diffuse fields in open systems and the emergence of the Green's function," J. Acoust. Soc. Am. **116**, 2731–2734 (2004).
[42]K. Wapenaar, E. Slob, and R. Snieder, "Unified Green's function retrieval by cross-correlation," Phys. Rev. Lett. **97**, 234301 (2006).
[43]O. A. Godin, "Emergence of the acoustic Green's function from thermal noise," J. Acoust. Soc. Am. **121**, EL96–EL102 (2007).
[44]R. L. Weaver and O. I. Lobkis, "Diffuse fields in ultrasonics and seismology," Geophysics **71**, SI5–SI9 (2006).
[45]K. Wapenaar, "Nonreciprocal Green's function retrieval by cross correlation," J. Acoust. Soc. Am. **120**, EL7–EL13 (2006).
[46]In structural acoustics, determination of wave fields in coupled fluid-solid systems is referred to as fluid-structure and flow-structure problems, depending on the presence of a background flow, while elastic waves within the solid are called structure-born sound.
[47]R. H. Mellen, "The thermal-noise limit in the detection of underwater acoustic signals," J. Acoust. Soc. Am. **24**, 478–480 (1952).
[48]T. Bowen, in *Acoustical Imaging*, edited by J. P. Powers (Plenum, New York, 1982), Vol. **11**, pp. 549–561.
[49]V. I. Passechnik, "Verification of the physical basis of acoustothermography," Ultrasonics **32**, 293–299 (1994).
[50]Yu. N. Barabanenkov and V. I. Pasechnik, "A study of correlation properties of thermal acoustic radiation," Acoust. Phys. **41**, 494–496 (1995).
[51]E. V. Krotov, M. V. Zhadobov, A. M. Reyman, G. P. Volkov, and V. P. Zharov, "Detection of thermal acoustic radiation from laser-heated deep tissue," Appl. Phys. Lett. **81**, 3918–3920 (2002).
[52]V. A. Vilkov, E. V. Krotov, A. D. Mansfel'd, and A. M. Reiman, "Application of focusing arrays to the problems of acoustic brightness thermometry," Acoust. Phys. **51**, 63–70 (2005).
[53]E. E. Godik, Yu. V. Gulyaev, Yu. V. Mirgorodskii, and V. I. Pasechnik, in *Acoustic Sensing and Probing: Fourth Course of the International School on Physical Acoustics* (World Scientific, Singapore, 1991), pp. 77–95.
[54]R. A. Kruger, D. R. Reinecke, and G. A. Kruger, "Thermoacoustic computed tomography—Technical considerations," Med. Phys. **26**, 1832–1837 (1999).
[55]L. D. Landau and E. M. Lifshits, *Statistical Physics*, Course of Theoretical Physics Vol. **5**, 3rd ed. (Reed, Oxford, 1980).
[56]O. A. Godin, "Reciprocity and energy theorems for waves in a compressible inhomogeneous moving fluid," Wave Motion **25**, 143–167 (1997).
[57]L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media. 2: Point Sources and Bounded Beams*, 2nd ed. (Springer, New York, 1999), Chap. 8.
[58]O. A. Godin, "Reciprocity relations and wave energy conservation for inhomogeneous flow/anisotropic solid systems," Acoust. Phys. **43**, 688–693 (1997).
[59]L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media. 1: Plane and Quasi-Plane Waves*, 2nd ed. (Springer, New York, 1998), Chap. 7.
[60]R. L. Weaver, "Ward identities and retrieval of Green's functions in the correlations of a diffuse field," Wave Motion **45**, 596–604 (2008).
[61]K. Aki and P. Richards, *Quantitative Seismology: Theory and Methods* (Freeman, San Francisco, 1980), Chap. 2.
[62]S. M. Rytov, "Correlation theory of thermal fluctuations in an isotropic medium," Sov. Phys. JETP **6**, 130–140 (1958).
[63]R. Snieder, K. Wapenaar, and U. Wegler, "Unified Green's function retrieval by cross-correlation; connection with energy principles," Phys. Rev. E **75**, 036103 (2007).
[64]P. Roux, K. G. Sabra, and W. A. Kuperman, "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84 (2005).
[65]R. Snieder, "Recovering the Green's function of the diffusion equation from the response to a random forcing," Phys. Rev. E **74**, 046620 (2006).
[66]H. B. Callen and T. A. Welton, "Irreversibility and generalized noise," Phys. Rev. **83**, 34–40 (1951).
[67]W. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, Cambridge, England, 1995), Chap. 3.
[68]O. A. Godin, D. Yu. Mikhin, and D. R. Palmer, "Monitoring ocean currents in the coastal zone," Izv. Akad. Nauk, Fiz. Atmos. Okeana **36**, 131–142 (2000).

Revisiting experimental methods for studies of acidity-dependent ocean sound absorption

Timothy F. Duda[a)]
AOPE Department, MS 11, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 28 October 2008; revised 5 February 2009; accepted 6 February 2009)

The practical usefulness of long-range acoustic measurements of ocean acidity-linked sound absorption is analyzed. There are two applications: Determining spatially-averaged *p*H via absorption measurement and verifying absorption effects in an area of known *p*H. The method is a differential-attenuation technique, with the difference taken across frequency. Measurement performance versus mean frequency and range is examined. It is found that frequencies below 500 Hz are optimal. These are lower than the frequency where the measurement would be most sensitive in the absence of noise and signal fluctuation (scintillation). However, attenuation serves to reduce signal-to-noise ratio with increasing distance and frequency, improving performance potential at lower frequencies. Use of low frequency allows longer paths to be used, with potentially better spatial averaging. Averaging intervals required for detection of fluctuations or trends with the required precision are computed. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3089591]

PACS number(s): 43.30.Pc, 43.30.Es, 43.30.Cq [RCG] Pages: 1971–1981

I. INTRODUCTION

Here, a multiple-frequency acoustic method to measure ocean *p*H-dependent sound absorption is reinvestigated. The method was first investigated by Jin and Worcester (1989). Using knowledge of how *p*H and absorption are related (Mellen and Browning, 1977; Brewer *et al.*, 1995), the method can be used either to determine spatially-averaged *p*H or to investigate details of absorption behavior if *p*H is known. Details regarding the first of these uses are the main focus of this paper, although the expressions can be modified for the reverse investigation. Applying the method as an acidity measure, ocean *p*H and temporal *p*H variations are estimated from the difference of sound attenuation at two (or more) frequencies. Being more specific, the technique would provide a nearly instantaneous remote measurement of sound absorption along an acoustic path, inferred mean *p*H along the path, and possibly a mean *p*H depth-profile in a slice filled by a set of acoustic paths oscillating vertically in the ocean acoustic waveguide. For studies of the temporal behavior of biogeochemical conditions over large physical scales, such an averaged *p*H measurement may (or may not) have sampling advantages over accumulated point measurements of *p*H in the situation of evolving ocean acidity having high-frequency or small-scale variability.

Global-mean ocean acidity must reflect future changes in atmospheric carbon dioxide concentration, which is rising, via flux of carbon dioxide into the ocean (see Doney *et al.*, 2009a). An estimate of the average surface ocean water *p*H reduction from preindustrial times to 1996 is 0.1 unit (Brewer, 1997; Orr *et al.*, 2005). This trend is predicted to continue, with further reduction by 0.3–0.5 units possible in the next 100 years (Caldeira and Wickett, 2005; Orr *et al.*, 2005) with consequences to biogeochemical systems (see Sec. II). Because sound absorption is inversely related to *p*H, it is possible that the ocean may become sonically noisier, all else constant (Hester *et al.*, 2008). Measurements of global *p*H conditions are thus desirable. If trends are spatially broad and temporally smooth, infrequent and sparse localized *p*H measurements would adequately detect the conditions. If there are significant small-scale variations as suggested by Doney *et al.* (2009b), then more frequent and comprehensive measurements would be required to detect trends amidst the fluctuations.

Measurements of average large-scale *p*H conditions are typically made by accumulating point measurements along transects (Fig. 1) via programs such as the World Ocean Circulation Experiment (WOCE), the Joint Global Ocean Flux Study, and follow-up studies, or with time series (Fig. 2). More frequent or spatially-averaged measurements could be made with the acoustic method, if required. An acoustic measure of *p*H would have less accuracy and precision than individual point measurements but may provide equally valid mean estimates in the case of considerable variability of the individual measurements.

The acoustic measurement methodology for *p*H-dependent absorption has been established by prior research (Jin and Worcester, 1989; Brewer *et al.*, 1995). However, thorough analysis and implementation have lagged theoretical and technical readiness. Recent advances in ocean acoustic systems and in observations of long-range signal fluctuations (which add uncertainty to the measurement) make this a good time to revisit this technology. Inexpensive and reliable sound projectors are available at frequencies of 200–600 Hz. These are suitable for long-term use at ocean observatories that are coming online, along with highly capable receiver systems assembled from off-the-shelf components. In light of this new technology and new interest in possible trends in ocean acidity, this paper builds on the

[a)]Electronic mail: tduda@whoi.edu

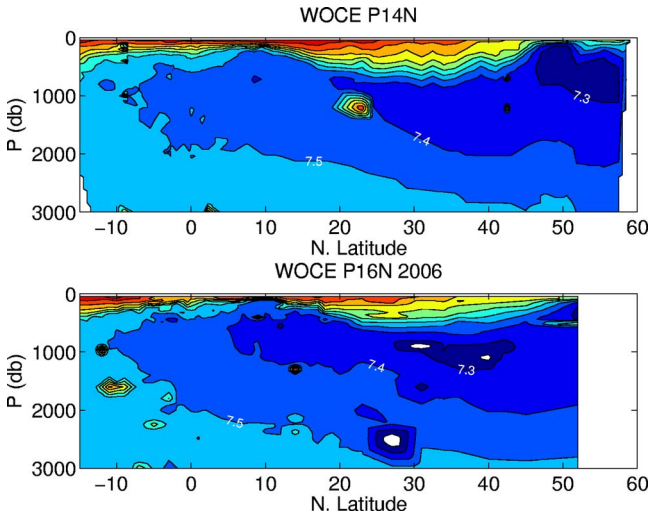


FIG. 1. (Color online) Full sections of *p*H in the tropical and North Pacific are shown. Data are from two WOCE transect cruises: P14 at longitude 180 in 1993 and P16 at longitude 150 W (approximately) in 2006. The contour interval is 0.1. The 7.3, 7.4, and 7.5 contours are labeled. Shading is similar to Fig. 2. Data are from the Carbon Dioxide Information Analysis Center (CDIAC, 2009a).

analysis of Jin and Worcester (1989) by investigating the measurement capability at many frequencies, bandwidths, and propagation distances. The mean acoustic energy differ-

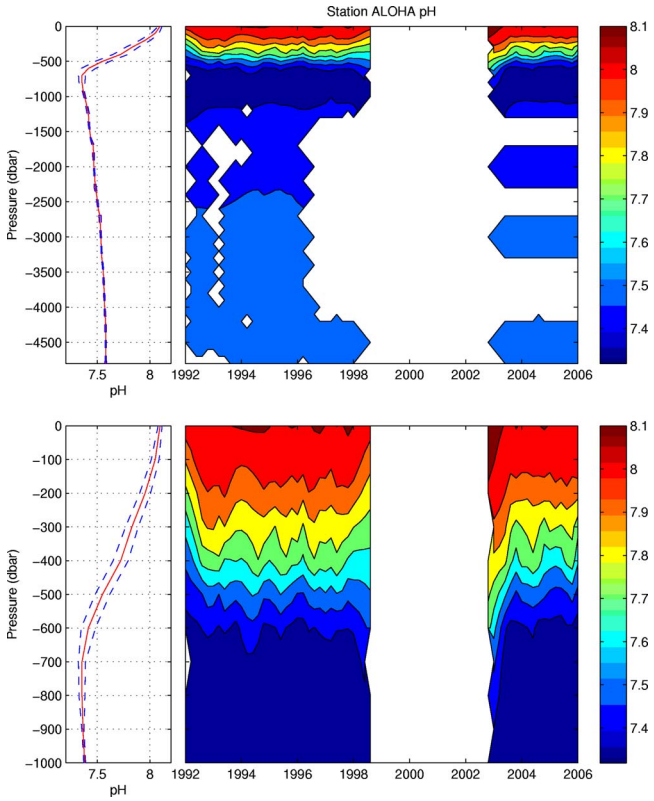


FIG. 2. (Color online) HOT Station ALOHA (2245′N, 158 W) *p*H data collected over a 14 year period. Data are from University of Hawaii (2009), tabulated and plotted by Bruce Howe. The data are collected several times per year. At the left are the time-mean profiles plotted as solid lines, with plus/minus standard deviation shown with dotted lines. To the right the time series are contoured, contour interval 0.1, with year indicated at bottom. The lower panels show an expanded view of the near-surface data, and the upper panels show all depths.

ence of two (or more) frequencies must be determined after propagation over fairly long distances of a few hundred kilometers, but such propagation degrades signal-to-noise ratio and introduces variability via scattering from mesoscale, internal wave, and surface wave phenomena. In this paper, signal averaging requirements necessary to achieve useful results are analyzed.

The methodology can be described very simply. Research has established the relationship between aqueous boric acid-borate equilibrium and energy absorption of sound below 1 kHz (Fisher and Simmons, 1977), although some anomalous behavior has been identified which suggests some interaction between borate and other ions occurs in the ocean (Mellen *et al.*, 1981, 1983). Pressure-sensitive reactions of the relevant ions modify the phase lag between pressure and velocity in sound waves, leading to energy loss. This is linked to *p*H because the concentration of hydroxide OH^- (Mellen and Browning, 1977) and/or carbonate CO_3^{2-} (Mellen *et al.*, 1983) plays a role. The kinetic behavior of the slowest reaction introduces a time scale of about a 100 ms, giving the 1 kHz transition frequency (Zeebe *et al.*, 2001). At high frequencies sound absorption is dominated by other processes and is not strongly dependent on *p*H, whereas below 10 kHz it is strongly *p*H dependent. This makes the shape of the absorption curve in the transition region, centered at about 600 Hz, *p*H-dependent in a manner that allows *p*H to be estimated from the shape of the curve. The absolute absorption at low frequency is also *p*H dependent, but can be difficult to measure because of uncertain net scattering loss. The shape of the absorption curve, determined by comparing signal level at two or more frequencies, is more sensible to study experimentally than absolute absorption because many propagation and scattering processes also affecting signal level are weak functions of frequency. That is, differencing over frequency minimizes detrimental scattering effects if they are weak functions of frequency (Jin and Worcester, 1989).

The differential technique can reduce or eliminate interference from mesoscale and weak internal-wave scattering, as long as such scattering does not cause wavefronts to break up into discernable micro-multi-path arrivals. However, some internal-wave induced fluctuations will exist, caused by interfering micro-fronts, and must be averaged over. Jin and Worcester (1989) recognized this as a major concern. They estimated the number of independent samples needed to provide useful measurements of *p*H (better than 0.05 accuracy) to lie in the range of 44–100 for the case of 450–650 Hz signals propagated 750 km, with sample independence attained when sample spacing exceeds the acoustic fluctuation time scale of tens of minutes. Results from internal-wave induced fluctuation research over the past two decades are incorporated here into a new analysis to re-determine the best frequencies to use and the number of samples needed. It is found that slightly lower frequencies are better, and that many more than the previous estimate of 100 samples may be needed.

The methodology is reviewed here from beginning to end, and the sampling requirements re-examined. First, motivation for ocean acidity and sound absorption studies is

discussed (Sec. II). Next, the physical chemistry of the problem is reviewed (Sec. III). In Sec. IV, the sensing method using long-range propagation is outlined [Jin and Worcester (1989) redux]. Section V examines the effectiveness of various frequency bands and path lengths in an idealized situation with steady acoustic propagation conditions and additive noise. Section VI shows how the probable presence of signal fluctuations dramatically increases the number of required samples that would need to be averaged together to obtain good results. Section VII summarizes the work.

II. MOTIVATION

Global ocean *p*H is likely to decrease because of ocean uptake of carbon dioxide deposited by society into the atmosphere. It is accepted that global surface *p*H has been reduced from 8.2 to 8.1 in the past 200 years (Brewer, 1997). Models predict that the surface *p*H will continue to drop as the atmospheric carbon dioxide concentration increases (Caldeira and Wickett, 2005). Local conditions such as vertical *p*H profiles throughout the world may be less amenable to prediction. A notable implication is that probable future inorganic carbon uptake by the ocean and oceanic *p*H reductions have the potential to profoundly affect calcium carbonate shelled or skeleton-producing animals, including (but not limited to) small pteropods, coral, clams, and lobsters (e.g., Feely *et al.*, 2004; Orr *et al.*, 2005, and references cited therein).

Changes in ocean *p*H whether from anthropogenic CO_2 or from other causes are equally important acoustically. The acid level responds to other factors besides uptake of carbon dioxide, such as decreased ventilation from changes in physical climate, many of which are also likely to reduce *p*H (Hester *et al.*, 2008). A simplified view of the carbon uptake effect is given: Aqueous carbon dioxide combines with water to form carbonic acid, which can dissociate into bicarbonate and hydrogen ions, or carbonate and two hydrogen ions. Because bicarbonate is very soluble, hydrogen ions and available carbonate ions will tend to form bicarbonate, which buffers the system against *p*H drop. However, this reduces carbonate concentration, which has the detrimental effect on calcium carbonate mineral structures. Even in water remaining supersaturated with respect to calcium carbonate, decreased calcification rates are expected (Orr *et al.*, 2005).

Such *p*H and carbonate changes would affect localized ecosystems at coastal margins (Feely *et al.*, 2008; Wootton *et al.*, 2008) as well as pelagic ecosystems and their fisheries such as tuna. Coastal ecosystem alteration would have significant natural and economic impact, and may indeed be closely tied to what is happening offshore. Therefore, both coastal and pelagic situations are of interest. The deep-sea ocean volume in some sense sets the boundary condition for regional acidity, and may serve as a reservoir or be a driver of change. Orr *et al.* (2005) and Feely *et al.* (2004) showed that subpolar regions are likely to be the first to reach undersaturation of calcium carbonate minerals calcite and aragonite, with probable pronounced biological impacts in those important areas.

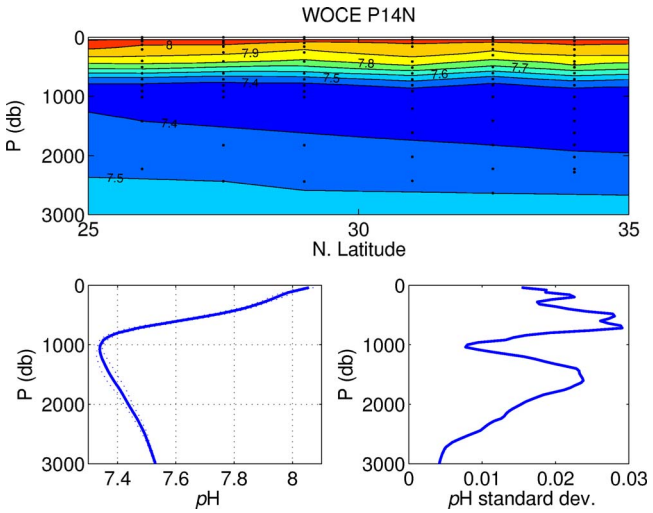


FIG. 3. (Color online) A portion of WOCE transect P14N from 1993 (Fig. 1) is shown in the top panel. The longitude is approximately 180°. The mean profile is shown at lower left, and the standard deviation profile at lower right. The dots in the upper panel show the locations of the data points. Data are from the Carbon Dioxide Information Analysis Center (CDIAC, 2009a).

Tracking *p*H in certain areas over time with effective averaging of short-term and small-scale variability may be desirable to assess the progress of these changes. However, representative measures of average deep-sea *p*H using ships are expensive. Sections of oceanic *p*H show mesoscale spatial and temporal variations (Fig. 1), which must be averaged over time and space in order to determine mean values and trends. At least two papers show scatter plots of oceanic *p*H from hydrographic cruises with fluctuations of ± 0.05 about a mean profile (Brewer *et al.*, 1995; Ternon *et al.*, 2001) illustrating the natural variability and possible sampling problem. Hawaii Ocean Timeseries (HOT) measurements show similar variability (Fig. 2), as do transects of *p*H (Fig. 3).

There are many features of the global carbonate system that may or not be stable. An example is the subsurface tongue of low *p*H seen in Figs. 1 and 3. This is associated with low surface *p*H in the subpolar region associated with carbon dioxide flux into the ocean at those higher latitudes due in part to thermodynamic effects (Takahashi *et al.*, 2006). The feature is consistent with low alkalinity tongues descending from the subpolar North Pacific in hydrography sections (CDIAC, 2009b). Similar features occur in the South Atlantic and South Pacific but not in the North Atlantic.

A full examination of *p*H variability in the Atlantic Ocean is available (Ternon *et al.*, 2001). The work shows that *p*H can be predicted from various local parameters (temperature, salinity, oxygen concentration, nitrate concentration, etc., not including carbon dioxide and carbonate) but only to within roughly 0.02.

Note that reduced ocean *p*H would mean reduced absorption of sound below a few kilohertz, possibly causing the ocean to become noisier at those frequencies, all else equal (Hester *et al.*, 2008).

One approach toward quantifying trends in oceanic *p*H, and associated sound absorption, and testing our ability to predict them would be implementation of a frequency-

dependent acoustic absorption system in the deep sea [Jin and Worcester (1989) type]. An acoustic system could automatically and make repeated instantaneous spatially-averaged measurements and may be able to make the absorption and *p*H measurements with useful precision and accuracy. This technique could prove cost effective with respect to ship-survey *p*H methods, particularly if implemented south of 40° S, one of the subpolar areas at risk of profound carbonate biogeochemistry change. Other reasons for measuring the *p*H of the pelagic ocean are that pelagic *p*H modeling may be more robust than in coastal regions and may be at the point of testability, and that spatial averaging over the known scales of natural variability with long-range acoustics is straightforward. The acoustic measurement resolution limit for *p*H will be shown to be about 0.05, larger than the variations shown in Figs. 2 and 3 for the central North Pacific, and thus would be unable to improve upon transect-mean estimates from areas having comparable variability. The acoustic technique would be most useful in areas with higher spatial variability or strong temporal variability.

III. *p*H -DEPENDENT SOUND ABSORPTION FROM BORATE RELAXATION

A solution containing the borate ion ($B(OH)_4^-$) and boric acid ($B(OH)_3$), present in small quantities in ocean water, changes volume when the dissolved boric acid/borate changes between three states, the end members being boric acid and borate. The volume change causes absorption as described below. Many coupled reactions may occur in concert with the borate reactions. A concise recent summary of uncertainties regarding the chemistry appears in Hester *et al.*, 2008. The simplest candidate reaction that may result in sound absorption is

$$B(OH)_3 + OH^- \leftrightarrow B(OH)_3 \cdot OH^- \leftrightarrow B(OH)_4^- \quad (1)$$

(Mellen and Browning, 1977; Mellen *et al.*, 1980), which does not involve other coupled reactions. This describes absorption data qualitatively but not quantitatively. Boric acid concentration exceeds borate at oceanic *p*H. This process roughly explains the frequency dependence of absorption because the right-hand step has the slower time constant than the other (somewhere around a millisecond) and yields a reduction in volume. The equilibrium state of Eq. (1) is pressure-dependent, explaining why acoustic waves trigger the transient volume changes. The system thus has what might be loosely described as a dynamic compressibility, which gives sound-speed an imaginary component (Liebermann, 1949). The borate acoustic absorption with relaxation frequency in the 1 kHz range was identified in the laboratory (Yeager *et al.*, 1973). That work followed earlier work involving salts with higher relaxation frequencies, primarily magnesium sulfate. Another system of reactions with similar absorption behavior is

$$\begin{array}{ccc} H_2O + B(OH)_3 + OH^- + CO_3^{2-} & \leftrightarrow & B(OH)_4^- + HCO_3^- + OH^- \\ \updownarrow & & \updownarrow \\ H_2O + B(OH)_3 \cdot OH^- + CO_3^{2-} & \leftrightarrow & B(OH)_4^- + H_2O + CO_3^{2-} \end{array} \quad (2)$$

(Mellen *et al.*, 1983). This is a reduced notation for a few possible systems that include pairing of some of the ions with metallic ions. The horizontal steps are the slow ones associated with low-frequency absorption. This system was proposed to account for absorption differences between solutions of borate alone and artificial seawater solutions containing dissolved inorganic carbon. Both Eqs. (1) and (2) involve *p*H via the role of the OH^- concentration.

The dissolution of CO_2 in the ocean and the subsequent formation of carbonic acid (H_2CO_3) change the concentration of hydroxide (OH^-) by releasing protons and forming three ions: bicarbonate (HCO_3^-), carbonate (CO_3^{2-}), and hydronium (H_3O^+, equivalent for our purposes to protons, H^+). (This changes the *p*H, the negative of the base-10 logarithm of hydronium concentration.) The hydronium ions react with the OH^- ions in the alkaline ocean to form water, changing the OH^- concentration, and thus also changing the relative concentrations of boric acid and borate, which alters the absorption because the transient reactions of Eq. (1) and/or Eq. (2) proceed at levels determined by the concentrations. In the weakly alkaline ocean, the reagent in Eq. (1) or Eq. (2) with least concentration is the hydroxide ion, so the reaction is closely coupled to its concentration. As a result, absorption is an increasing function of *p*H (decreasing function of acidity). This *p*H dependency is independent of the sum concentration of borate and boric acid, which is essentially proportional to salinity.

Two additional chemical species cause absorption, magnesium sulfate and magnesium carbonate, but they have higher relaxation frequencies. Their absorptions are weak with respect to that of borate and are essentially constant with respect to frequency and *p*H at the low frequencies of interest here.

IV. DIFFERENTIAL-ATTENUATION MEASUREMENTS OF ABSORPTION

Using notation in Brewer *et al.* (1995), the coefficient of sound absorption (units dB/km) can be written as

$$\alpha(f) = \alpha_b(f) + \alpha_m(f) + \alpha_c(f). \quad (3)$$

Below 1 kHz the *p*H-sensitive borate term α_b dominates. Below 700 Hz the other terms (magnesium sulfate α_m and magnesium carbonate α_c) can be neglected (Fisher and Simmons, 1977). Signal power in decibels is defined as a ratio $P = 10\log_{10}(I/I_0)$, where I is intensity (peak amplitude squared for a sine wave) and I_0 is a reference value. The borate term can be written as

$$\alpha_b = (S/35)A_b(f^2 f_b/(f^2 + f_b^2)), \quad (4)$$

where $A_b = 0.1$ K, $f_b = 0.9 \times 10^{T/70}$ kHz, S is salinity (psu), and T is temperature (°C). The *p*H-dependence (NBS *p*H scale) is contained in $K = 10^{(p\mathrm{H}-8)}$. Figure 4 shows borate absorption versus frequency for various *p*H values at fixed salinity and temperature. The measurement of absorption (and thus of *p*H if the other parameters are sufficiently well known) cannot be reliably made by measuring signal strength at one frequency in the ocean because acoustic forward scattering processes from mesoscale eddies, surface en-

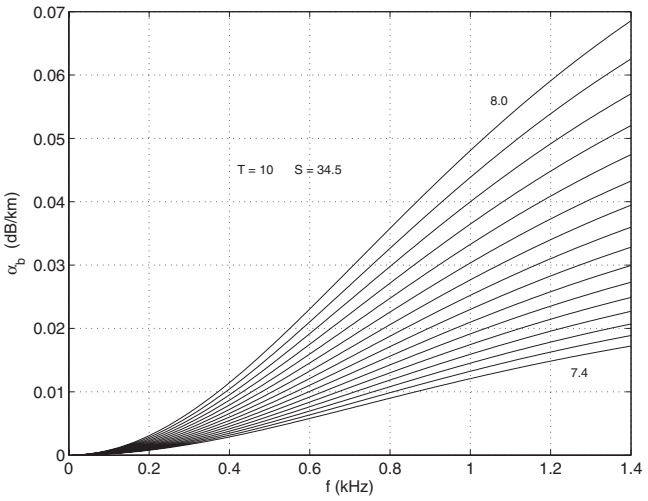


FIG. 4. Borate absorption α_b as a function of frequency is plotted for temperature 10° and salinity 34.5 for *p*H from 7.4 to 8.0 by 0.04 increments. Total absorption values at 1 kHz are 18% higher (*p*H 8.1) to 65% higher (*p*H 7.4) than the borate absorption.

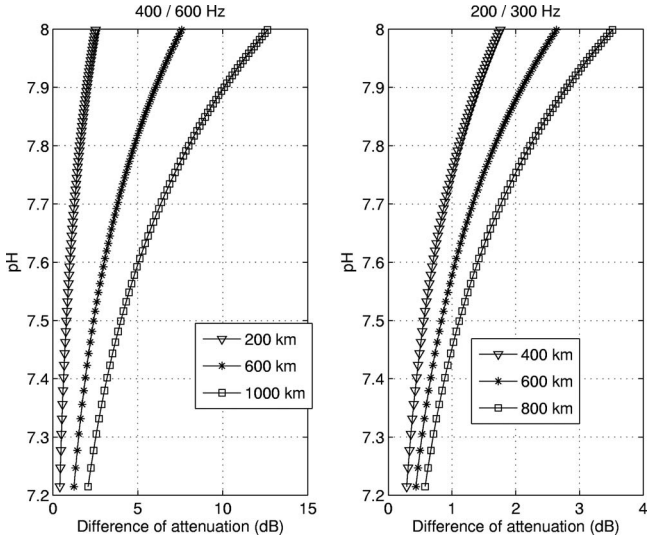


FIG. 5. D_A versus *p*H computed using expression (10) is plotted for two different frequency pairs and a few propagation distances.

counter, and seabottom encounter cause large fluctuations in energy. However, *p*H-dependent sound absorption due to borate is a function of frequency, so that the fingerprint of this process can be obtained using energy measurement at two or more frequencies, enabling recovery of *p*H. This assumes that scattering loss differences between the frequencies are small. For the frequencies investigated here, this is most questionable for surface reflecting rays in high sea states (Erbe and Farmer, 2000; Brekhovskikh and Lysanov, 1982). Such rays may be the only type available for long-range paths in polar regions; in other areas fully refracted rays should be available.

Loosely following Jin and Worcester (1989) but altering the equations to include updated information in Brewer *et al.* (1995), the differential attenuation along a single propagation path Γ with arc increment ds is given by

$$D_A(f_1,f_2) = \int_\Gamma \alpha(f_2)ds - \int_\Gamma \alpha(f_1)ds = \int_\Gamma \int (\partial\alpha(s)/\partial f)dfds, \quad (5)$$

where f is frequency in kilohertz. The paths are the same, so after retaining only the dominant term α_b,

$$D_A = \int_\Gamma (S(s)/35)A_b(s)f_b(s)\left(\frac{f_2^2}{f_2^2+f_b^2(s)} - \frac{f_1^2}{f_1^2+f_b^2(s)}\right)ds. \quad (6)$$

The position-dependent terms (those dependent on s) are functions of either T or *p*H. If *p*H is known then measurement of D_A would test the veracity of the expressions for K, A_b, and f_b. Alternatively, comprehensive inversion of a group of ray-integrals (6) using methods in Munk *et al.*, 1995 would provide S, K (thus *p*H), and f_b (thus T) as functions of position in a source-receiver vertical slice. However, acoustic data alone would probably be insufficient for a complete two-dimensional inversion. A reasonable approach more likely to yield reliable inversion would be to divide the integrals into a few parts each corresponding to depth increments, such as 0–100 m, 100–200 m, etc., and to incorporate auxiliary information to refine the solution (such as limiting *p*H fluctuations to the upper kilometer, for instance). Temperature structure (affecting f_b) is sufficiently well-known to allow inversion for *p*H. A test-case published by Jin and Worcester (1989) derived a solution for a path that could be approximated with a constant T equal to 4°. This is the temperature near the sound channel axis at lower latitude sites (such as HOT, Fig. 2). A constant low temperature assumption is also suitable for polar regions. At this temperature, the frequency f_b=1 kHz (note that f_b increases to 2 kHz at 24°), and integration gives the approximate solution

$$D_A = A_bLC \quad (7)$$

for S equal to 35, where L is source-receiver distance in kilometers, and where

$$C = \frac{f_2^2}{f_2^2+1} - \frac{f_1^2}{f_1^2+1}. \quad (8)$$

From this, using the definition $A_b = 0.0237\times 10^{(0.83p\mathrm{H}-6)}$, they obtain

$$p\mathrm{H} = 0.83^{-1}\left[6 + \log_{10}\left(\frac{D_A}{0.0237LC}\right)\right]. \quad (9)$$

Using our definition of A_b and letting S vary gives the slightly different expression

$$p\mathrm{H} = 8 + \log_{10}\left(\frac{350D_A}{SLC}\right). \quad (10)$$

Figure 5 shows the differential absorption that must be measured at various frequencies and ranges to determine *p*H using this constant temperature approximation [i.e., using Eq. (10)]. Salinity of 35 psu is assumed. The measurement sensitivity is given by $d(pH)/dD_A = (\ln(10)D_A)^{-1} = 0.4343/D_A$. From Fig. 5, at 600 km range, bandwidth of 200 Hz (400, 600 Hz), a ΔpH of 0.05 at *p*H 7.6 yields ΔD_A of 0.58 dB.

Source and receiver depth placement can contribute to how Eq. (6) is utilized to investigate differential attenuation and/or *p*H. Use of near-axial acoustic rays to determine *p*H, which travel through water where *p*H and T are relatively constant, as opposed to vertically cycling rays which integrate over vertically varying T, *p*H, and thus α_b, is the scenario leading to Eqs. (7) and (10), but those paths undergo much scattering and may have amplitude fluctuations that make them unusable (Duda *et al.*, 1992; Colosi *et al.*, 1994) (effects of scattering are discussed in Sec. VI). Polar regions also have near-constant temperature (a few °C variations), although the rays cycle vertically and reflect from the surface, which may add frequency-dependent surface loss. A full vertically-dependent inverse determination would be straightforward to implement and would show a slightly different relationship between distance, D_A, and mean *p*H than Eq. (10), but Eq. (10) is used as a guideline for measurement sensitivity in Sec. V, as done by Jin and Worcester (1989). Note that receivers at multiple depths in the oceanic sound channel on a single mooring would receive eigenray signals that may be spaced such that their internal-wave fluctuation effects are independent, adding reliability to the measurement. Section VI addresses how independent measurements must be used to build statistical reliability of the method.

In a full inversion, a depth profile of *p*H is a likely result. This procedure is analogous to temperature inversion, and would require differential attenuation to be measured along more than one acoustic eigenray. Simplification (i.e., depth function parametrization) may be helpful if the approximate depth profile of *p*H is known and has its shape constrained by known processes, or if the depth profile of rms *p*H fluctuations is known (i.e., fluctuations limited to the near-surface region).

The crucial quantity α_b is T- and S-dependent at fixed *p*H, as shown. In situations where this constant T approximation is usable, neglect of expected natural T and S variations along the acoustic path would lead to errors of not more than 4% in forward modeling of D_A. This would give errors smaller than 0.03 in *p*H computation from observed D_A.

V. MEASUREMENT EFFECTIVENESS VERSUS FREQUENCY AND RANGE

To satisfy the goal of making meaningful continuous spatially-averaged measurements, long measurement ranges are desired to average over mesoscale features. The primary acoustic data, D_A, grow with range and become more easily measurable. Thus, there are two reasons why long propagation paths are desirable. This would imply that low frequencies are necessary because of the prototypical frequency dependence of absorption, which is empirically in the Pacific, for example (disregarding details appearing in Sec. IV) (Munk *et al.*, 1995)

$$\alpha(f) = 0.79A\frac{f^2}{0.8^2+f^2} + \frac{36f^2}{5000+f^2} \quad \text{dB/km}, \tag{11}$$

where f again has units of kilohertz and the geographically variable quantity A is 0.055 for the North Pacific. On the other hand, the measurement sensitivity to frequency, $|\partial\alpha/\partial f|$, related to D_A, has a peak near 0.6 kHz (Fig. 6), apparently diminishing the effectiveness of lower frequencies, as reasoned by Jin and Worcester (1989). This is shown here to not be the entire story, with a balance occurring between low differential attenuation at low frequency and short range, and low signal strength at high frequency and long range.

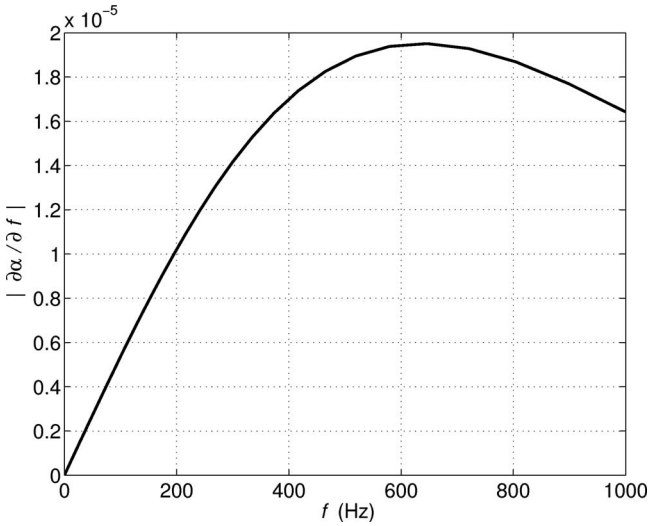


FIG. 6. The derivative of the absorption curve is shown.

A. Signal method

The system design assumed in this performance analysis section is one capable of making acoustic measurements of discrete eigenray arrivals (wavefront segments, equivalently) that are common to two (or more) frequency bands. This can involve looking at the band-edges of a wideband signal (as with Jin and Worcester, 1989), which would have lower signal to noise than the band center, or it can involve looking at two independent bands from either one source or two sources. The amplitude difference is taken to obtain D_A. In addition to occupying two or more bands, the signals in each band must have enough bandwidth to provide good coherent processing gain via pulse-compression techniques (Munk *et al.*, 1995). This is required to improve signal-to-noise ratio and to identify acoustic ray geometry if full inversion of Eq. (6) is desired. Signal strength measurements (i.e., attenuation) would be made using peak amplitudes of pulse-compression output for isolated identified eigenrays. Alternatively, spectral level of the pulse-compression output for the isolated arrivals could be used, which would correspond to a base-band version of the received signal spectrum with significant signal power gain, with gain given by the ratio of the total signal duration divided by the inverse signal bandwidth (or code digit length). Ideally, signals in each band would be transmitted at the same time from very closely spaced sources. The best situation would be to use one source with a very broad bandwidth, more than 300 Hz.

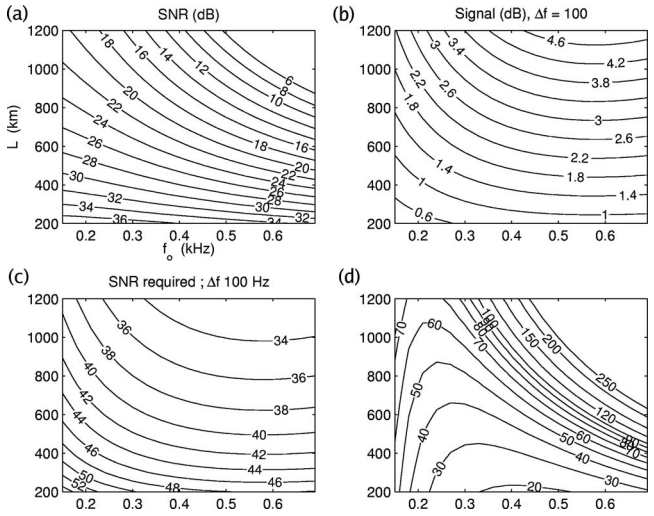


FIG. 7. Four plots display contours of four different quantities versus center frequency and path length L of a dual-frequency differential-attenuation system. (a) SNR from formula (12) is shown. PG=24 dB, SL=185 dB. (b) Signal level D_A for 100 Hz frequency difference, pH=7.8 is shown. (c) SNRR to achieve the D_A-scaled ε requirement [from (b)] for detection of a pH difference of 0.05. This is computed from Eq. (13) after determining the ε requirements from the D_A curves in (b). (d) The number of samples needed to achieve the required ε (or SNRR) computed from information in panels (a) and (c) using Eq. (14b).

B. Dependency of signal-to-noise ratio

To determine the best frequency (without building an expression for optimum frequency at this time), begin with the sonar equation

$$\mathrm{SNR}(f,L) = \mathrm{SL} - \mathrm{TLa}(f,L) - \mathrm{TLg}(L) - (\mathrm{NL}(f) - \mathrm{PG}). \tag{12}$$

The terms are signal-to-noise ratio (SNR), source level (SL), absorption transmission loss (TLa), geometrical spreading transmission loss (TLg), noise level (NL), and processing gain (PG) [PG is substituted for the conventional array gain (AG)]. Here, set source level=185 dB, noise level=65 dB (moderate sea state), and PG=24 dB for a coded pulse. TLg is conservatively expressed as spherical spreading, which approximately accounts for the division of energy into multiple arrivals by having excess loss beyond that of cylindrical spreading that might better describe pure tone transmission. Figure 7(a) shows SNR for these parameters as a function of frequency and range. SNR decreases with frequency and range.

C. Dependency of differential attenuation

Figure 7(b) shows $D_A(f,L)$ from Eq. (7) modeled for 100 Hz separation. The signal D_A increases with range and center frequency, except above 600 Hz. Note the SNR and the signal (D_A) have opposed trends in terms of suitability with respect to frequency and range.

An important aspect of the inversion problem is the accuracy of D_A measurement. For oceanic transmission measurements, there are two effects which cause difficulty in measuring this difference of two mean signal levels. One is the fact that the measurement is composed of signal plus noise, not just signal. The other is the frequency-dependence of natural signal fluctuations caused by interfering microfronts.

In the presence of Gaussian noise, Jin and Worcester (1989) showed that upper-bound of measurement error for signal strength, in decibels, is given by

$$\varepsilon = 20\log_{10}(1 + \sqrt{2} \times 10^{-\mathrm{SNRR}/20}), \tag{13}$$

where SNRR is the realized signal-to-noise ratio after pulse compression and averaging, also in decibels. The raw SNR can be improved using short-term signal averaging. Averaging N measurements gives realized signal-to-noise ratio of $\mathrm{SNRR}=\mathrm{SNR}+10\log_{10} N$. For a desired ε of 0.1 dB and a reasonable estimated SNR of 26 dB, for example, close to 40 measurements would be needed to achieve the required SNRR of 42 dB.

In the inverse sense of the problem, the required ε for an effective measurement of pH from D_A can be computed as a function of frequency and range. Expression (13) can be used to obtain SNRR, and SNR and SNRR can be used to obtain N. A starting point is the derivative of Eq. (10) given earlier. From the 200–300 Hz D_A curve in Fig. 5 for L=600 km and pH=7.8, mean D_A is about 1.7 dB and $\Delta D_A = \Delta p\mathrm{H} D_A/0.4343$ is equal to 0.20 dB for ΔpH=0.05. To measure this, assume that signal level at each frequency must be determined to an accuracy of $\Delta D_A/4$ (i.e., ε is required to be less than 0.05 dB in each band for center frequency 250 Hz, L=600 km). (These requirements are rough, and can be improved during design of an actual system using statistical methods. They serve the purpose here, yielding approximate measurement requirements versus frequency and range. Absolute numbers of required measurements may vary slightly.) Because D_A is a function of range and frequency, the required $\varepsilon(f,L)/\varepsilon(0.25,600) = DA(f,L)/DA(0.25,600)$. Figure 7(c) contours $\mathrm{SNRR}(f,L)$ computed directly from these $\varepsilon(f,L)$ using Eq. (13) and the scaling requirements obtained from data of Fig. 5.

Figure 7(d) shows the required number of samples $N(f,L)$ needed for a successful detection of delta-pH of 0.05, computed from SNRR (Fig. 7(c)) and SNR (Fig. 7(a)). Two formulas for this are

$$\sqrt{N} = (\sqrt{2} \times 10^{-\mathrm{SNR}/20})/(10^{\varepsilon/20} - 1), \tag{14a}$$

$$N = 10^{(\mathrm{SNRR-SNR})/10}. \tag{14b}$$

There is a "trough" in the plot showing a low N requirement extending from [400 Hz, 200 km] to [250 Hz, 800 km] (lower center to upper left). The measurement is possible at any of the ranges and frequencies covered in the figure, but the fewest samples are needed at the shortest ranges where SNR is high. If chemical measurement issues demand that longer ranges be used, frequencies of 250–400 Hz would be best, as indicated by the trough.

To summarize this section, for the inverse problem (pH determination), the balance between the undesirable decrease in SNR with frequency and range, and the desirable DA improvement with increasing range and frequency has been quantified. This is independent of the desirability to propagate long distances to achieve good spatial averaging.

VI. INTERNAL-WAVE EFFECTS

Next, frequency-dependent fluctuations of the arrivals that are caused by internal waves are considered. The differencing would reject fluctuations common across the measurement band, but that is a best-case scenario. The worst case is treated here. Many of the fluctuation effects caused by internal waves are somewhat coherent across frequency, but they are treated here as independent at the two frequencies. For fluctuations with normal statistics, the mean-squared error of the mean estimate is equal to the variance divided by the number of samples, MSME$=\mu_2/N$. This allows fluctuation parameter statistics to be converted to estimates of the quality of sampled means. The presence of internal-wave induced intensity fluctuations is commonly known, and numerous theoretical and computational studies exist. However, the intensity fluctuation variance at 500 km range is poorly delineated in the literature. This is because difficulties such as inadequate SNR have left comparisons of data with scattering theory unable to fully test the theories.

It is known that an impulsive sound wavefront in the temperate ocean sound channel deforms into an accordion shape (in space at fixed time, and in time at a fixed distance from the source) (Munk *et al.*, 1995). Each leg of the accordion corresponds to a single member of the multi-path arrival structure and to a specific number of upper turning points of the looping sound rays plus an initial direction index (two choices, up or down). Each arrival at a point receiver corresponds to sound that is concentrated along eigenrays connecting source to receiver. The intensity and phase fluctuations of each arrival can be estimated using integrals of ocean properties such as sound-speed anomaly along each ray (Flatté *et al.*, 1979; Flatté and Rovner, 2000). Many of the effects causing intensity (I) fluctuation, such as fluctuating focus effects along the rays due to internal waves, are weakly dependent on frequency. Thus, D_A would be only weakly sensitive to them. In the ideal case of infinite SNR and frequency independence of internal-wave effects, a borate absorption measurement could be made with a single measurement of intensity difference at two frequencies, taking only an instant.

However, intensity fluctuations exist which vary across a frequency band on an instantaneous basis. Experiment and computer simulations have shown that internal waves can cause each of the arrivals to be composed of many micro-multi-path arrivals (Colosi *et al.*, 1994; Simmen *et al.*, 1997). Interference between these arrivals can cause intensity fluctuations that are stronger functions of frequency than the intensity fluctuations of intact wavefronts because interference is a function of phase, and phase in each band of broadband sound propagated over a fixed path is a function of frequency. The result is that intensity can fluctuate rapidly and independently at two measurement frequencies. Thus, a scintillation in the intensity difference over the two frequencies (which is related to our desired parameter D_A via $D_A = 10\log_{10}(I_2/I_1)$) will result.

Jin and Worcester (1989), using a spectral formalism, computed that about 100 intensity-difference measurements would be needed to suitably average the intensity fluctuations in order to achieve a measurement error of 0.3 dB for 550 Hz, range 750 km in the North Pacific. The expression is $\varepsilon = 10\log_{10}(1+((M-1)/2NM)^{1/2})$, where M is the number of interfering micro-rays and N is the number of averaged receptions. More averaging would be required to achieve 0.1 dB (see Sec. V C).

There are a few published numbers for the total intensity fluctuation. The scintillation index is defined to be the intensity variance divided by the square of the mean intensity, $SI=\mu_2/\mu_1^2$. This has been observed to be between 0.2 and 2.0 for various resolved early arrivals in a 1000 km experiment, with an average value of 0.9 (Duda and Bowlin, 1994). These numbers include macro effects of internal waves that are weak functions of frequency, and thus overestimate the fluctuations that should remain after frequency differencing. Values far above 1.0 correspond to strong interference of a few signals, those at 1.0 correspond to saturation (interference of many arrivals), and values less than 0.3 correspond to weak scattering (Born or Rytov) (Flatté *et al.*, 1979). Note that intensity statistics are not normal, but can be approximated as such for the smaller observed values of SI. This is not a severe problem because this is a study of confidence intervals and error bars, and the improvements gained via averaging will be realized nonetheless, although improvement is more difficult to quantify than in the case of normally distributed fluctuations.

It is convenient to obtain a unified expression for the number of measurements needed to obtain the required precision in the presence of the two noise processes: internal-wave induced intensity fluctuation and additive Gaussian noise. Define signal waveform as $S(t)=S_0(t)+S_{IW}(t)+n(t)$, where S_0 is mean signal, S_{IW} is additive zero-mean internal-wave fluctuation, and n is noise. Using methods in Jin and Worcester (1989) and expressions from Sec. V, one can write for the typical measurement error at each frequency

$$\varepsilon = 20\log_{10}\left(1+\frac{\langle|S_{IW}|^2\rangle^{1/2}}{\langle|S_0|^2\rangle^{1/2}}+\sqrt{2}\times 10^{-(\mathrm{SNR}+10\log N)/20}\right). \tag{15}$$

The triangle inequality has been used, so this is an upper bound on the deviation between the mean of measured energy S^2 and the desired energy S_0^2 [i.e., on $10\log_{10}(\langle S^2\rangle/\langle S_0^2\rangle)$]. Now assume that the second term in brackets, the fluctuation amplitude contribution, is reduced with averaging. With normal statistics, the second term becomes $SA/\sqrt{N}$. Our newly defined scintillation amplitude statistic $SA=\langle|S_{IW}^2|\rangle^{1/2}/\langle|S^2|\rangle^{1/2}$ is (oddly) similar in shape to SI for many reasonable pdf's of S_{IW}, but there is no simple analytic relationship. For fluctuations with a normal distribution, SA/SI exceeds 1 for low SI but tends to have a value near 0.62 for SI greater than 0.7. SA is about $SI/2$ for exponentially distributed populations with SI near 1.4. Rearrangement of Eq. (15) yields

$$\sqrt{N}=(SA+\sqrt{2}\times 10^{-\mathrm{SNR}/20})/(10^{\varepsilon/20}-1). \tag{16}$$

In the limit of SA going to zero, this expression properly reverts to Eq. (14a) which applies to the case of steady signal with noise.

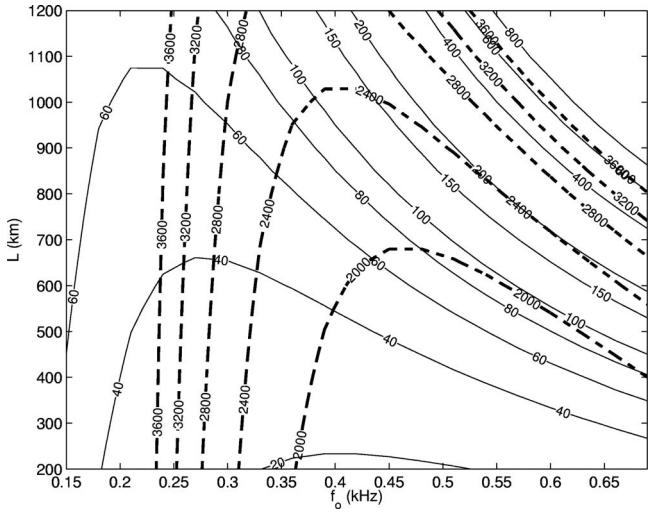


FIG. 8. The required number of samples for *p*H measurement with 0.05 precision for the situation of fluctuating signal and noise, and 100 Hz frequency difference, is contoured over center frequency and range (bold dashed lines). The same quantity for the steady signal and noise situation is also shown (solid lines, repeated from Fig. 5(d)).

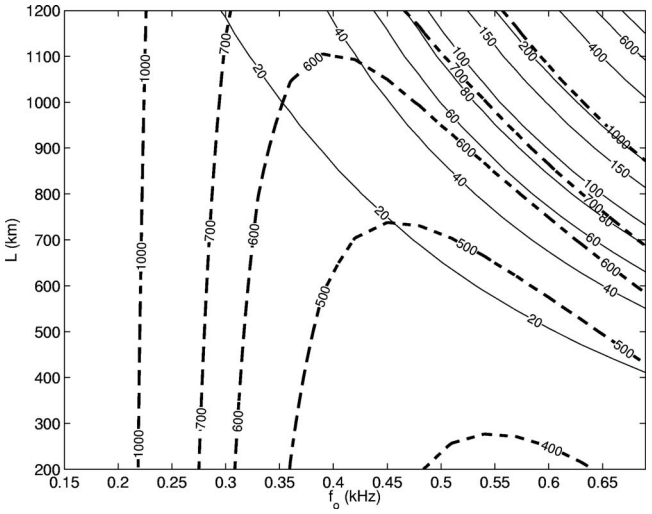


FIG. 9. The number of samples needed to measure *p*H to an accuracy of 0.05 units using a 200 Hz bandwidth measurement is contoured. Dashed contours are for the case that includes signal fluctuations and noise, and the solid contours are for the case of noise only.

The second term of the numerator is typically about 0.11 (22 dB SNR), whereas the scintillation term [for individual wavefront (ray) arrivals] is likely to always exceed this. This means that the signal fluctuations will compromise the measurement ability much more than noise. This agrees with the comparison of the two effects made by Jin and Worcester (1989). The most stringent scenario that they calculate is that 225 measurements would be needed to average over signal fluctuations to reduce error to 0.2 dB. Figure 8 contours the number of samples needed for realistic values of *SA* and shows that many more than their estimates may be needed. Modeling *SA* to be linear with range, starting at $SA=0$ at zero range and growing to $SA=1.0$ at 1000 km range, the plot shows contours of *N* needed to reduce ε to the value needed to effectively sample the expected *DA* to a *p*H precision of 0.05 (analogous to procedures of the prior section). These are calculated from Eq. (16). Some of the *N* contours for the noise-only result of Fig. 7(d) are replotted. The two sets of contours have different shapes, and the fluctuating signal result shows that far more measurements must be averaged than in the noise-only case. There is an order of magnitude difference at range of 800 km with 600 Hz. For the case that includes fluctuations, there is a "trough" of greatest effectiveness (minimum number of points) passing through 400 Hz, 1000 km. If *SA* is reduced by 50% (i.e., weaker internal-wave scattering is assumed), the *N* values fall to about one-third of the values indicated in the contour plot over the area of the trough, which retains its shape. Figure 9 shows results computed in a similar fashion to those of Fig. 8 except a band separation of 200 Hz is used. The time required to obtain the required order-500 samples can be estimated. For a fluctuation coherence time of 5 min, the time would be about 2 days. As with the 100 Hz separation situation shown in Fig. 8, reducing SA by 50% decreases *N* by about 70% with little change in the shape of the contours.

There is a possibility that internal-wave induced amplitude fluctuations of distinct arrivals are correlated over frequency. In this case it would be advantageous to difference the peak amplitudes of the pulses before averaging (predifferencing). Noise would remain, but fluctuation effects might be reduced or eliminated. An expression for difference ε can be developed. With noise treated as independent over frequency, the expression will look like Eq. (15) except a difference amplitude statistic can be substituted for *SA*. Because the differencing is done first, the tolerance for ε for a given delta-*p*H is eased with respect to that used earlier by a factor of 2 (i.e., the value is 0.12 dB for delta-*p*H of 0.05 at 600 km, 250 Hz). The amplitude statistic [second term in Eq. (15)] can take the form $A=\langle|S_{IW1}|-|S_{IW2}|\rangle/\langle|S_{01}|-|S_{02}|\rangle$, where subscripts indicate frequency band one or two. Averaging then yields $SA\sqrt{2/N}$ for the case of uncorrelated fluctuations. This method produces curves similar to those of Fig. 8, but approximately double the numbers of samples indicated in that figure are needed. Thus, adding uncorrelated fluctuations before the averaging process increases their detrimental effect. However, if internal-wave fluctuations are correlated in the two bands, then very improved performance would result. For perfect correlation of internal-wave fluctuations across the band, the noise-only results would hold for the pre-difference technique.

VII. SUMMARY

The technique of using differential acoustic attenuation measurements at frequencies below 600 Hz to determine the precise nature of absorption effects (direct measurement) and by extension average *p*H (inverse measurement) has been examined. The basic processes are reviewed and the equations are derived. The capabilities of many frequency difference pairs and various acoustic path lengths at providing attenuation measurements of the required accuracy in a noisy environment are examined.

Differential attenuation (energy difference between two frequency bands) might be measurable with equipment of existing types, either using a broad sweep system (Morozov

and Webb, 2003) or two systems, each with bandwidth of 50 Hz or more, operating in bands separated by 100 or 200 Hz (or more). Using established formulas, estimates are made of the required measurement accuracy, in decibels, to make meaningful *p*H measurement (inverse problem) for various propagation distances, various center frequencies, and two-frequency band separations. The results are in the form of contours of the quantity of independent samples required to be averaged together to measure effects consistent with *p*H fluctuations of 0.05.

Two scenarios are examined: steady signal (idealized) and fluctuating signal (more realistic). Separated multi-path signal peaks used to estimate energy received after propagating distances of a few hundred kilometers are not steady, but fluctuate, largely due to internal-wave effects. The fluctuations of energy in the field have not been well characterized in the literature. These fluctuations are of order 1 in comparison with the mean energy, so in the high SNR case they dominate the error of the measurement. Conservative estimates of the number of samples required are made assuming the fluctuations are independent in the two frequency bands whose energy is differenced to obtain the measurement. The values are a few thousand for 100 Hz separation, and order 500 for 200 Hz separation.

The results reflect a competition between two effects. Attenuation difference grows with range and has a maximum at about 600 Hz. Signal-to-noise ratio shrinks with range and frequency. The result is that paths of 600–1000 km and frequencies of 375–450 Hz may be most practical for large-scale sensing. These frequencies are below the frequency of maximum sensitivity (600 Hz), and their effectiveness stems from having better SNR than 600 Hz. Using three rather than two frequencies would provide a slightly better measure of the attenuation curve than would a simple two-frequency difference by adding information.

To confront small open questions regarding the chemistry (Mellen *et al.*, 1983; Zeebe *et al.*, 2001) and the dependencies in Eq. (4), differential-attenuation measurements of absorption can be combined with ship surveys to verify absorption behavior. Regarding the inverse problem (*p*H measurement), the method would provide a spatially-averaged measure that is less precise than individual *in situ* *p*H measurements. It would be most useful in two applications. One is in regions of high spatial variability, where unrealistically large numbers of *p*H measurement profiles would be required to obtain a good mean estimate. The other application is a time series measurement of *p*H variability. This may be most cost-effective in remote regions where a time series involving *in situ* samples would be difficult. A few hundred transmissions through independent internal-wave fields (a source of noise in the attenuation measurement), possible in a period of a few days, would provide a mean *p*H value. Repeat transmission cycles provide the time series. Time series could also be provided by moored autonomous *p*H sensors, but fixed-site point sensors may be at risk of not representing large-scale mean *p*H conditions very well because of local bias effects. Clearly there are advantages and disadvantages of each measurement technology. Perhaps the strongest potential for the acoustic measure is in remote polar areas, where undersaturation of calcium carbonate minerals may occur first, with strong biological consequences.

ACKNOWLEDGMENTS

Funding was provided by Woods Hole Oceanographic Institution. Discussions with Bruce Howe and Peter Worcester were helpful. Following suggestions made by the editor and Peter Brewer has significantly improved the manuscript.

Brekhovskikh, L., and Lysanov, Yu. (**1982**). *Fundamentals of Ocean Acoustics* (Springer-Verlag, Berlin).

Brewer, P. (**1997**). "Ocean chemistry of the fossil fuel CO_2 signal: The haline signal of "Business as Usual"," Geophys. Res. Lett. **24**, 1367–1369.

Brewer, P. G., Glover, D. M., Goyet, C., and Shafer, D. K. (**1995**). "The *p*H of the North Atlantic Ocean: Improvements to the global model for sound absorption," J. Geophys. Res. **100**, 8761–8776.

Caldeira, K., and Wickett, M. E. (**2005**). "Ocean model predictions of chemistry changes from carbon dioxide emissions to the atmosphere and ocean," J. Geophys. Res. **110**, C09S04.

CDIAC (**2009a**). Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, U.S. Department of Energy, http://cdiac.ornl.gov/oceans/pacific.html (Last viewed January, **2009**).

CDIAC (**2009b**). Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, U.S. Department of Energy, http://cdiac.ornl.gov/oceans/glodap/Glopatlas.htm (Last viewed January, **2009**).

Colosi, J. A., Flatté, S. M., and Bracher, C. (**1994**). "Internal-wave effects on 1000-km ocean acoustic pulse propagation: Simulation and comparison to experiment," J. Acoust. Soc. Am. **96**, 452–468.

Doney, S. C., Lima, I., Feely, R. A., Glover, D. M., Lindsay, K., Mahowald, N., Moore, J. K., and Wanninkhof, R. (**2009b**). "Mechanisms governing interannual variability in upper-ocean inorganic carbon system and air-sea CO_2 fluxes: Physical climate and atmospheric dust," Deep-Sea Res., Part II (to be published).

Doney, S. C., Tilbrook, B., Roy, S., Metzl, N., Le Quéré, C., Hood, M., Feely, R. A., and Bakker, D. (**2009a**). "Surface ocean CO_2 variability and vulnerability," Deep-Sea Res., Part II (to be published).

Duda, T. F., and Bowlin, J. B. (**1994**). "Ray-acoustic caustic formation and timing effects from ocean sound-speed relative curvature," J. Acoust. Soc. Am. **96**, 1033–1046.

Duda, T. F., Flatté, S. M., Colosi, J. A., Cornuelle, B. D., Hildebrand, J. A., Hodgkiss, W. S., Jr., Worcester, P. F., Howe, B. M., Mercer, J. A., and Spindel, R. C. (**1992**). "Measured wave-front fluctuations in 1000-km pulse propagation in the Pacific Ocean," J. Acoust. Soc. Am. **92**, 939–955.

Erbe, C., and Farmer, D. M. (**2000**). "A software model to estimate zones of impact on marine mammals around anthropogenic noise," J. Acoust. Soc. Am. **108**, 1327–1331.

Feely, R. A., Sabine, C. L., Hernandez-Ayon, J. M., Ianson, D., and Hales, B. (**2008**). "Evidence for upwelling of corrosive "acidified" water onto the continental shelf," Science **320**, 1490–1492.

Feely, R. A., Sabine, C. L., Lee, K., Berelson, W., Kleypas, J., Fabry, V. J., and Millero, F. J. (**2004**). "Impact of anthropogenic CO_2 on the $CaCO_3$ system in the oceans," Nature (London) **305**, 362–366.

Fisher, F. H., and Simmons, V. P. (**1977**). "Sound absorption in sea water," J. Acoust. Soc. Am. **62**, 558–564.

Flatté, S. M., Dashen, R., Munk, W. H., Watson, K. M., and Zachariasen, F. (**1979**). *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, London).

Flatté, S. M., and Rovner, G. (**2000**). "Calculations of internal-wave-induced fluctuations in ocean-acoustic propagation," J. Acoust. Soc. Am. **108**, 526–534.

Hester, K. C., Peltzer, E. T., Kirkwood, W. J., and Brewer, P. G. (**2008**). "Unanticipated consequences of ocean acidification: A noisier ocean at lower *p*H," Geophys. Res. Lett. **35**, L19601.

Jin, G., and Worcester, P. F. (**1989**). "The feasibility of measuring ocean *p*H by long-range acoustics," J. Geophys. Res. **94**, 4749–4756.

Liebermann, L. (**1949**). "Sound propagation in chemically active media," Phys. Rev. **76**, 1520–1524.

Mellen, R. H., and Browning, D. G. (**1977**). "Variability of low-frequency sound absorption in the ocean: *p*H dependence," J. Acoust. Soc. Am. **61**, 704–706.

Mellen, R. H., Browning, D. G., and Simmons, V. P. (**1981**). "Investigation of chemical sound absorption in sea water: Part III," J. Acoust. Soc. Am. **70**, 143–148.

Mellen, R. H., Browning, D. G., and Simmons, V. P. (**1983**). "Investigation of chemical sound absorption in sea water: Part IV," J. Acoust. Soc. Am. **74**, 987–993.

Mellen, R. H., Simmons, V. P., and Browning, D. G. (**1980**). "Low-frequency sound absorption in sea water: A borate-complex relaxation," J. Acoust. Soc. Am. **67**, 341–342.

Morozov, A. K., and Webb, D. C. (**2003**). "A sound projector for acoustic tomography and global ocean monitoring," IEEE J. Oceanic Eng. **28**, 174–185.

Munk, W., Worcester, P., and Wunsch, C. (**1995**). *Ocean Acoustic Tomography* (Cambridge University Press, New York).

Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F., Yamanaka, Y., and Yool, A. (**2005**). "Anthopogenic ocean acidification over the twenty-first century and its impact on calcifying organisms," Nature (London) **437**, 681–686.

Simmen, J., Flatté, S. M., and Wang, G.-Y. (**1997**). "Wavefront folding, chaos, and diffraction for sound propagation through ocean internal waves," J. Acoust. Soc. Am. **102**, 239–255.

Takahashi, T., Sutherland, S. C., Feely, R. A., and Wanninkhof, R. (**2006**). "Decadal change of the surface water pCO_2 in the North Pacific: A synthesis of 35 years of observations," J. Geophys. Res. **111**, C07S05.

Ternon, J. F., Oudot, C., Gourlaouen, V., and Diverres, D. (**2001**). "The determination of pH_T in the equatorial Atlantic Ocean and its role in the sound absorption modeling in seawater," J. Mar. Syst. **30**, 67–87.

University of Hawaii (**2009**). School of Ocean an Earth Science and Technology, Hawaii Ocean Timeseries, http://hahana.soest.hawaii.edu/hot (Last viewed January, **2009**).

Wootton, J. T., Pfister, C. A., and Forester, J. D. (**2008**). "Dynamic patterns and ecological impacts of declining ocean pH in a high-resolution multi-year dataset," Proc. Natl. Acad. Sci. U.S.A. **105**, 18848–18853.

Yeager, E., Fisher, F. H., Miceli, J., and Bressel, R. (**1973**). "Origin of the low-frequency sound absorption in sea water," J. Acoust. Soc. Am. **53**, 1705–1707.

Zeebe, R. E., Sanyal, A., Ortiz, J. D., and Wolf-Gladrow, D. A. (**2001**). "A theoretical study of the boric acid-borate equilibrium in seawater," Mar. Chem. **73**, 113–124.

Estimating cetacean population density using fixed passive acoustic sensors: An example with Blainville's beaked whales

Tiago A. Marques[a] and Len Thomas
Centre for Research into Ecological and Environmental Modelling, The Observatory, University of St Andrews, St Andrews KY16 9LZ, Scotland

Jessica Ward and Nancy DiMarzio
Naval Undersea Warfare Center Division, 1176 Howell Street, Newport, Rhode Island 02841

Peter L. Tyack
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 2 December 2008; revised 3 February 2009; accepted 5 February 2009)

Methods are developed for estimating the size/density of cetacean populations using data from a set of fixed passive acoustic sensors. The methods convert the number of detected acoustic cues into animal density by accounting for (i) the probability of detecting cues, (ii) the rate at which animals produce cues, and (iii) the proportion of false positive detections. Additional information is often required for estimation of these quantities, for example, from an acoustic tag applied to a sample of animals. Methods are illustrated with a case study: estimation of Blainville's beaked whale density over a 6 day period in spring 2005, using an 82 hydrophone wide-baseline array located in the Tongue of the Ocean, Bahamas. To estimate the required quantities, additional data are used from digital acoustic tags, attached to five whales over 21 deep dives, where cues recorded on some of the dives are associated with those received on the fixed hydrophones. Estimated density was 25.3 or 22.5 animals/1000 km^2, depending on assumptions about false positive detections, with 95% confidence intervals 17.3–36.9 and 15.4–32.9. These methods are potentially applicable to a wide variety of marine and terrestrial species that are hard to survey using conventional visual methods. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3089590]

PACS number(s): 43.30.Sf, 43.80.Ka [WWA] Pages: 1982–1994

I. INTRODUCTION

Cetaceans (whales and dolphins) form a key part of marine ecosystems, and yet many species are potentially threatened with extinction by human activities. One essential element of an effective conservation or management strategy is a reliable estimate of population size ("abundance") or, equivalently, number per unit area ("density"). However, most cetacean species are hard to survey, since they live at low density over large areas of ocean and spend almost all of their time underwater. The object of this paper is to increase the repertoire of tools available for making species assessments, by developing and demonstrating methods for estimating cetacean density from surveys of their vocalizations collected from fixed passive acoustic sensors.

Currently, the main method for obtaining estimates of density is through visual line transect surveys. A set of randomly placed lines is traversed by an observation platform (e.g., ship, airplane, or helicopter) and all sighted animals of the target species are recorded, together with their perpendicular distance from the line. In the standard method, it is assumed that all animals on the transect line (i.e., at zero distance) are seen with certainty, but that probability of detection declines with increasing distance from the line. The distribution of observed detection distances is then used to estimate the average probability of detection, and this in turn allows estimation of population abundance or density. Line transects are a special case of distance sampling methods, which are described in detail in the two standard texts by Buckland *et al.* (2001, 2004).

Visual line transect methods have a number of disadvantages for surveying cetaceans: they can only be performed during daylight hours and are strongly dependent on good weather conditions; they do not work well for species that spend long periods of time underwater; they are expensive to do well and have restricted temporal coverage. On the other hand, some cetacean species make frequent and characteristic vocalizations, and this has led to increasing recent interest in the use of passive acoustic methods for monitoring cetacean populations (see review by Mellinger *et al.*, 2007b). One solution is to replace or supplement the visual observers on a shipboard line transect survey with a towed passive acoustic platform, since even a simple two-element hydrophone array can be used to obtain locations of repeatedly vocalizing animals, and hence the required perpendicular distances. This has proved particularly effective for sperm whales (*Physeter macrocephalus*), which are long, deep divers and hence hard to detect visually, but produce loud

[a]Electronic mail: tiago@mcs.st-and.ac.uk; URL: http://www.creem.st-and.ac.uk/decaf/

 0001-4966/2009/125(4)/1982/13/$25.00

echolocation clicks (Hastie *et al.*, 2003; Barlow and Taylor, 2005; Lewis *et al.*, 2007).

Acoustic line transect methods still require an expensive survey platform and have restricted temporal coverage. By contrast, technology exists where fixed acoustic sensors (either anchored or buoyed) can be deployed to record data over long periods or transmit it to shore via cables. Mellinger *et al.* (2007b) provide a review of the potential modalities. Such sensors have the potential to provide relatively inexpensive long-term monitoring and, if deployed in a wide-baseline array, could also cover large spatial areas.

There are several examples of the use of fixed acoustic sensors to provide an index of abundance, or at least to detect the presence of a species over time (e.g., Au *et al.*, 2000; Charif *et al.*, 2001; Moore *et al.*, 2006; Mellinger *et al.*, 2007a; Stafford *et al.*, 2007). However estimating absolute density or abundance from such data is considerably more difficult, although considerably more useful too. Broadly, the key problems are as follows: (i) to convert the number of vocalizations detected to the number produced within a defined area, by accounting for false positive and false negative detections, and (ii) to convert this estimate of vocalizations produced to the average number of animals present. To date, no method has been proposed that might be used in general to estimate cetacean density from fixed passive acoustic devices (Mellinger *et al.*, 2007b), although several papers have addressed elements of the problem (e.g., Aubauer *et al.*, 2000; McDonald and Fox, 1999; Wiggins *et al.*, 2004; Stafford *et al.*, 2007; Zimmer *et al.*, 2008).

This paper presents a framework for estimating cetacean density from an array of fixed passive acoustic detectors. The framework lays its foundation in a type of distance sampling survey methodology called cue counting (Buckland *et al.*, 2001) and is general enough that it might be used under considerably different scenarios, with appropriate modifications that are also discussed. The proposed approach is illustrated by a case study: estimating the density of Blainville's beaked whale, *Mesoplodon densirostris* (Md), at the Atlantic Undersea Test and Evaluation Center (AUTEC) range, an instrumented US Navy testing range in the Bahamas. Part of the data used here were also used in a previous attempt to estimate Md density at AUTEC by Moretti *et al.* (2006), which provides a useful comparison for the results obtained.

II. STATISTICAL FRAMEWORK FOR ESTIMATING DENSITY

Density D is defined as a number of animals per unit area, and hence density can be expressed as

$$D = N/a, \tag{1}$$

where N is the number of animals present in a region of area a. It is therefore straightforward to convert any density estimate into an abundance estimate, as long as the area it applies to is known, as $N=Da$. Given the number of animals n detected in survey area a, an intuitive estimator of abundance is given by

$$\hat{D} = \frac{n}{a\hat{P}}, \tag{2}$$

where $\hat{P}$ represents the estimated probability of detecting an animal. P can be estimated by a number of methods, of which distance sampling (Buckland *et al.*, 2001) is arguably the most commonly used.

A. Proposed estimator and variance

Here it is assumed that a set of replicate fixed passive acoustic sensors has been deployed using a random design (e.g., systematic random) within the survey region of interest (see, for example, Buckland *et al.*, 2001, Chap. 7 and Buckland *et al.*, 2004, Chap. 7, for discussions of appropriate sampling designs). The goal is to estimate average density and/or abundance of a cetacean species over the time period that the sensors were deployed. The methods generalize easily to multiple species, spatial and temporal stratification, etc. They are also applicable to single sensor situations and nonrandom sensor placement, although in these cases caveats must be introduced.

The methods proposed here involve counting (for a known period of time) the number of detected acoustic cues produced by the animals of interest, and appropriately scaling up this number of detected cues to estimate animal density. The appropriate definition of a cue is study-specific, but it will usually be a distinct animal vocalization such as a song unit, echolocation click, etc. Considering Eq. (2), an estimate of density $\hat{D}$ from the n_c detected cues over a time period T (and additional information detailed below) can be obtained by

$$\hat{D} = \frac{n_c(1-\hat{c})}{K\pi w^2 \hat{P} T \hat{r}}, \tag{3}$$

where w is the distance away from the hydrophones beyond which cues are assumed to not be detected, $\hat{P}$ is the estimated average probability of detecting a cue made within distance w, $\hat{r}$ is the estimated cue production rate, $\hat{c}$ is the estimated proportion of false positive detections, and K is the number of replicate sensors used.

Note that Eq. (3) can be seen as a special case of Eq. (2). $n_c(1-\hat{c})$ corresponds to the number of detected cues that were actually from the target species. $T\hat{r}$ corresponds to the number of cues produced by an average animal during the recording time T. Hence $n_c(1-\hat{c})/T\hat{r}$ corresponds to n in Eq. (2), while $K\pi w^2$ corresponds to a in Eq. (2).

Strictly speaking, the hydrophones should be independent sampling units, but in practice some clicks are detected at more than one hydrophone. However, Buckland (2006) showed by simulation that these methods are insensitive to this sort of nonindependence.

Assuming independence of the various random components, the variance for the density estimator can be approximated via the delta method (Seber, 1982) (see also Powell, 2007, for an applied review of the method) as

$$\mathrm{var}(\hat{D}) \approx D^2\{CV(n_c)^2 + CV(\hat{r})^2 + CV(\hat{c})^2 + CV(\hat{P})^2\}, \quad (4)$$

where CV denotes the coefficient of variation of the estimate, i.e., the standard error of the estimate divided by the estimate.

The variance in n_c can be estimated from the empirical variance of the cue counts n_{cj} $(j=1,2,\ldots,K)$ over K hydrophones, as is usually done for conventional distance sampling methods (Buckland *et al.*, 2001, p. 78) as

$$\mathrm{var}(n_c) = K\sum_{j=1}^{K} \frac{[n_{cj} - n_c/K]^2}{K-1}.$$

For the other three components, obtaining the appropriate variance will depend on how these are estimated. Examples are given in the case study.

Confidence intervals on density can be obtained by assuming that the density estimate follows a log-normal distribution (Buckland *et al.*, 2001, p. 77). An alternative approach for obtaining both variance and confidence intervals is the nonparametric bootstrap [Buckland *et al.*, 2001; see section 3.6 for details in this context, and Efron and Tibshirani, 1993 for an overview]. Both the analytic and bootstrap methods are commonly employed in conventional distance sampling.

Sections II B–II D consider methods for obtaining estimates of the three components P (probability of detection), r (cue production rate), and c (false positive proportion).

B. Probability of detection

To understand the potential methods for estimating P, it is convenient to start by describing the estimation of the detection function, $g(y)$, that is the probability of detecting a cue given it is produced at horizontal distance y. Here y refers to horizontal distance (i.e., projected onto the sea surface); the use of direct (i.e., slant) distances is discussed later.

In conventional distance sampling methods, including cue counting, $g(y)$ is estimated using the horizontal distances of detected objects from the line or point (usually referred to as "transect," with the equivalent here being a point with the acoustic sensor at the center, from which acoustic cues are detected). This is feasible because, with random transect placement, the *true* distribution of object distances is known; hence any departure in the distribution of *detected* distances from this known distribution can be attributed to the detection process. Specifically, in variants of distance sampling where the transect is a point, as in the current case, the true distribution of horizontal object distances, $h(y)$, is given by $2y/w^2$, where w is some distance beyond which detections are not recorded (and can be infinity). Thus, the probability density function (pdf) of observed distances, $f(y)$, is given by

$$f(y) = \frac{h(y)g(y)}{\int_0^w h(y)g(y)dy} = \frac{2yg(y)}{w^2 P}, \quad (5)$$

where $P = \int_0^w h(y)g(y)dy$ acts as a normalizing constant that ensures $f(y)$ is a pdf (i.e., that integrates to 1). Note that $g(y)$ can be multiplied by any constant without changing $f(y)$; in other words, the observed distances give information about the shape of $g(y)$ (i.e., the change in detectability with distance) but not the absolute level of $g(y)$. Therefore, an additional assumption must be made before $g(y)$ can be estimated, and the conventional assumption is that all objects at zero distance are detected, i.e., that $g(0)=1$. Then, given a parametric model for $g(y)$, the parameters of the model can be estimated from a sample of distances using, for example, maximum likelihood estimation.

Once an estimate of the detection function parameters is obtained, it is straightforward to estimate the average probability of detecting an object within w, as the average of the estimated detection probability at each distance multiplied by the density of objects at that distance:

$$\hat{P} = \int_0^w \hat{g}(y)h(y)dy. \quad (6)$$

Specialized software, DISTANCE (Thomas *et al.*, 2006), is freely available for the analysis of such data, and variance estimates and confidence limits are readily available from the output.

The conventional approach is sometimes possible using data from fixed acoustic sensors, so long as (i) horizontal distances to detected cues can be obtained, and (ii) the basic assumptions are met: that the true distribution of horizontal object distances, $h(y)$, is known and that $g(0)=1$. Calculation of horizontal distances using a single bottom-mounted sensor was demonstrated by McDonald and Fox (1999), using transmission loss and multi-path methods to estimate distance to calling fin whales near a Hawaiian seamount, and by Wiggins *et al.* (2004), using waveguide propagation to estimate distances to North Pacific right whales in the Bering Sea. However, the assumption of known $h(y)$ is often questionable with a single sensor. If cues are received on multiple sensors and can be associated, then time differences of arrival can potentially be used to localize the source, thereby providing a distance (e.g., Freitag and Tyack, 1993). However, the conventional methods would then need to be extended to allow the input data to be a set of detections of the same cue from multiple sensors at different distances.

An alternative way to obtain the detection function is to derive it from assumed models for sound production, propagation, and detection. This approach was demonstrated (using different modeling approaches and on different species) by Zimmer *et al.* (2008) and Stafford *et al.* (2007). Such approaches are naturally heavily dependent on the accuracy of the models and input parameters.

A third approach is to estimate the detection function empirically, using a set of experimental "trials" where vocalizations are produced at known times and distances from sensors, and the outcome is whether they are detected or not. One way to achieve this is to play recorded vocalizations, but a more realistic method involves placing acoustic recording tags on a sample of animals that are within the vicinity of the sensors, and using the tag data to determine where and when vocalizations were produced. Ideally, this additional experiment should be undertaken concurrently with the primary survey and on a representative sample of animals and sensors; otherwise differences in conditions between the tagging experiment and the primary survey will need to be accounted

for, by modeling covariates thought to influence detectability and then predicting detectability at the time of the trials. This is the approach taken in the case study, and the analysis options are described more fully later.

C. Cue production rate

The cue production rate r can be estimated using acoustic tag data, as described at the end of the Sec. II B, and this is the approach taken in the case study. Again, ideally, this would be collected concurrently with the primary survey and on a representative sample of animals. Otherwise, factors that potentially affect cue rate such as season, time of day, sex, etc., will need to be accounted for.

A second approach, if additional experimental tagging data are unavailable, is to use a cue rate estimate published elsewhere. For some species, which have very predictable vocalization patterns, this may produce a reliable estimate; for others it will not.

A third alternative, if it is possible to associate successive vocalizations from individual animals within the sensor array (such as when calls are uniquely identifiable or density is very low), and if it can be determined when animals are so close to a sensor that no calls will be missed, is to use the cue rate from vocalizations produced by these animals. However, for species that can pass by a sensor without producing any cues, this will inevitably lead to an overestimate of cue rate.

D. False positive proportion

To be recorded as a cue, a sound must go through a two-stage process: it must be (i) detected as being potentially relevant and (ii) classified as the correct type of sound. Because of the volume of data involved, both stages are typically performed by computer algorithms.

If a sound of interest, i.e., a vocalization by an animal from the target species within distance w of a sensor, is not detected, or is detected but incorrectly classified, it is referred to as a "false negative." (In statistics, this is also known as a type II error, considering the combined detection and classification process to be a statistical test.) The detection function, via the estimated detection probability, corrects for this type of error.

On the other hand, it is possible for a sound that is not a vocalization from the target species to be detected and classified as a cue, creating a "false positive" (or a type I error). The proportion of sounds classified as cues that do not come from the target species is the false positive proportion, and this must be estimated. As with the previous components, there are multiple ways to achieve this.

The most direct approach is to take a sample of the sounds classified as cues and manually check what proportion are false positives. This is the approach taken in the case study, and relies on the human operator to make correct classifications. If a human operator cannot be considered reliable, an alternative is to use some performance statistics of the detector/classifier, for example, on data where animals were known to be vocalizing because they were tagged or being observed from a nearby boat, and predict the average false positive proportion under the conditions of the survey being analyzed. This requires knowledge of all the variables that affect the false positive proportion, and their distribution both for the dataset where the performance characteristics were tested and for the dataset for which density is being estimated. This seems hard to implement as it implicitly requires one to know the total number of positive and negative sounds (i.e., whale clicks versus sounds from other sources). Hence the first approach is to be preferred wherever it is feasible.

Note in general, one cannot generalize to other scenarios the proportion of false positives in a given place and time. In particular, if the relative abundance of species contributing to false positives changes, the false positive proportion will change accordingly.

III. CASE STUDY: BLAINVILLE'S BEAKED WHALES AT AUTEC

The proposed methods are illustrated here with a detailed case study: the estimation of average density of Blainville's beaked whales, Md, at the AUTEC underwater tracking range during a 6 day period in spring 2005. The inference made is restricted to the time periods, during these 6 days, for which recordings were available (see below for details).

Beaked whales are considered difficult to survey using visual methods, since they make long, deep dives and are relatively difficult to sight even for the short periods of time they are at the surface (Barlow *et al.*, 2006; Barlow and Gisiner, 2006). However, they make regular high-frequency echolocation clicks during the deeper part of their dives (Johnson *et al.*, 2004) which, although highly directional (Zimmer *et al.*, 2005), can be detected by both surface-and bottom-mounted hydrophones, with the latter likely to have better detection ranges (Zimmer *et al.*, 2008). Md have previously been detected up to 6500 m slant range from the AUTEC hydrophones (Ward *et al.*, 2008).

AUTEC is an instrumented US Navy testing range located in the Tongue of the Ocean, Bahamas. It contains a wide-baseline array of 93 cabled hydrophones, mounted 4–5 m off the sea floor at depths between 1.3 and 2 km, and with an upward, roughly hemispherical beam pattern. For the data presented in this study, 82 hydrophones were digitally recorded at 96 kHz (Fig. 1): 68 wideband hydrophones spaced 4 km apart with a usable bandwidth from 50 Hz to 45 kHz, and 14 hydrophones spaced 1.8 km apart with a bandwidth from 8 to 50 kHz (bandwidth and spacing values are approximate). Previous research utilizing this hardware has demonstrated the presence of Md at AUTEC, developed suitable automated detectors and estimated approximate animal density (Moretti *et al.*, 2006; DiMarzio *et al.*, 2008; Ward *et al.*, 2008).

The processed data and metadata used in this paper were submitted to the Ocean Biogeographic Information System and are accessible via http://seamap.env.duke.edu/.

A. Case study data and data processing

Two sets of data were used in the case study. The first (the "primary dataset") was a set of recordings taken on 82

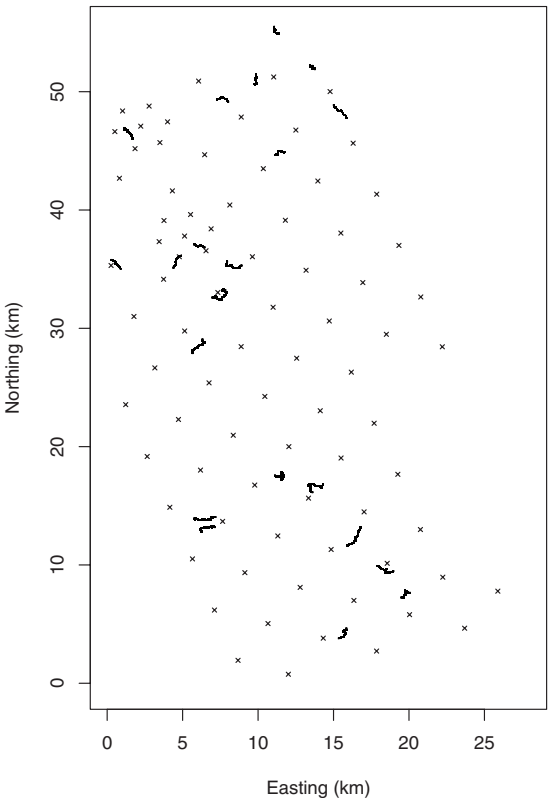


FIG. 1. The spatial layout of the AUTEC hydrophones that were recording during the collection of the primary survey data, represented by small crosses. Also shown as dots (perceived as solid lines) are the locations of the tagged Blainville's beaked whales when each click recorded on the DTag was produced.

hydrophones during 6 days in spring 2005 (between 17:03 GMT, 26 April and 08:32 GMT, 2 May 2005). These were processed to extract detections of Md foraging clicks (the cues), and the number of detections was used as the basis for estimating average density of Md over the 6 day period. For operational reasons the data were divided into 12 sets, and entire sets (2, 5, 7, and 11), as well as some scattered minutes, were not used. These periods were removed from the analysis for various reasons (e.g., problems with acoustic recordings, system tests being conducted over the range, etc.).

Traditional hyperbolic localization techniques used to obtain the position of sound sources such as whales require a click to be detected on at least four hydrophones. However, previous studies have observed that the combination of Md narrow beam width and AUTEC's wide hydrophone baselines result in clicks often not being detected on more than one or two hydrophones simultaneously (Ward *et al.* 2008); hence detection function and cue rate estimation methods based on localizing Md at AUTEC were considered infeasible. Therefore, a second dataset was required to obtain these quantities. This (the "auxiliary dataset") comes from acoustic tagging studies carried out in October 2006 and August–September 2007 in which five whales were fitted with digital acoustic tag (DTag) acoustic recording tags (Johnson and Tyack, 2003) which remained attached for a total of 21 deep dives (Table I). The tag data were used to estimate the cue rate. A subset of the data was further processed to estimate the location of the whales at each click event, and these clicks were matched with clicks received on surrounding AUTEC hydrophones to enable estimation of the detection function.

TABLE I. Details about the tagged whales used in case study analysis: Tag ID, date the animal was tagged, number of dives while animal was tagged, and number of dives with data available for estimating the detection function $g(y)$.

Tag	Date	Number of Dives	Dives for $g(y)$
Md296	23 Oct 2006	3	3
Md227	15 Aug 2007	6	0
Md245	2 Sep 2007	4	3
Md248a	5 Sep 2007	4	4
Md248b	5 Sep 2007	4	3
Total		21	13

A brief description follows of the processing required to extract detections from the main dataset, and to extract click rates, locations, and associations with AUTEC hydrophones from the auxiliary dataset.

1. Processing of primary dataset

Clicks were detected using a multi-stage fast Fourier transform (FFT)-based energy detector (Morrissey *et al.*, 2006), using data digitally recorded from the range hydrophones at a 96 kHz sampling rate. A 2048-point FFT with 50% overlap was used. This provides a 46.875 Hz frequency resolution and a 10.67 ms time resolution. During the first stage, an adaptive threshold was run in each bin of the FFT. If energy above threshold was detected, the bin was set to a "1" and a detection report was generated. During the second stage, detections are classified based on the frequency distribution of the detected signal. Detections are classified as beaked whale clicks by evaluating the ratio of bins above threshold to the total number of bins within the frequency band of 24–48 kHz (Morrissey *et al.*, 2006; Ward *et al.*, 2008). Additionally, a second check is performed by examining the number of bins set outside the 24–48 kHz "beaked whale band." If this exceeds 10%, then the detection is reclassified as a dolphin as they are more likely to have significant spectral energy below 24 kHz.

The detection reports classified as beaked whales were then summed into a final output consisting of the number of beaked whale clicks per minute for each hydrophone over the recording period.

2. Processing of auxiliary dataset

Additionally, DTag (Johnson and Tyack, 2003) data, both positional and acoustic, were available for five whales, in a total of 21 (Table I) deep dives. The tag records continuously from a built-in hydrophone and suite of sensors that convey information on positional data: animal depth, heading, pitch, and roll. The sensors sample the orientation of the animal in three dimensions with sufficient resolution to capture individual fluke strokes (Johnson and Tyack, 2003). The speed from the pressure sensor, heading, and pitch angle were used to reconstruct a three-dimensional "pseudotrack"

for each animal during the period the tag was attached (Johnson and Tyack, 2003). These data were used for the estimation of click rates, and also for simulation of click characteristics involved in the estimation of the mean detection probability (see below).

For the estimation of the detection function, further data processing was required. For four whales and a total of 13 of the deep dives, these DTag data were associated with matched filter detections from the surrounding hydrophones, localized, and the pseudotrack georeferenced (i.e., absolute rather than relative coordinates obtained; see Ward *et al.*, 2008 for further details). Once georeferenced, the DTag accelerometer and magnetometer data for each click within the georeferenced portion of the dive were used to calculate the spatial relationship between the whale and each hydrophone in the whale orientation frame (see Johnson and Tyack, 2003; Ward *et al.*, 2008, for further details). An indicator of detection (1—detected; 0—not detected) could then be generated for each click produced at each surrounding hydrophone (using the same FFT-based detector as in the primary survey click count data), as well as a set of corresponding click characteristics in the frame of the whale (namely, slant distance, vertical, and horizontal off-axis angles with respect to the whale's orientation).

B. Case study density estimation methods

1. Estimating probability of detection

As stated earlier, the conventional distance sampling approach of estimating average detection probability using detection distances could not be used due to the difficulty in easily localizing Md clicks using the wide-baseline hydrophone array at AUTEC. Hence auxiliary data from the tagging study were used, where for each click produced by a tagged whale, it was known which of the surrounding hydrophones detected the click.

If tags had been applied to a large random sample of animals from the population present at AUTEC during the 6 day period of interest, then the average proportion of clicks that were detected would be an unbiased estimator of the average detection probability required. However, the tags were applied in a different time period, and the sample size of tags was not large. For both of these reasons, a more complex approach was required, where the probability of detection was modeled as a function of covariates thought to influence detectability, and the fitted relationship then used to estimate average detection probability during the 6 day period.

The modeling approach used was a generalized additive model (GAM, Wood, 2006), with a binomial response and logistic link function, to model the probability of detecting a Md click as a function of available covariates. The dependent variable used was the outcome of the detection and classification process (i.e., detected and classified as a cue/not detected or not classified as a cue) for each click produced by the tagged whale at each hydrophone within 8 km of the location of the whale (at the moment of the first click in the corresponding dive). The explanatory variables considered were the slant whale-hydrophone distance and the horizontal (*hoa*) and vertical (*voa*) off-axis angles, measured on the frame of the whale, with respect to the hydrophone (e.g., a 0,0 angle pair corresponds to a whale pointing straight at the hydrophone). For slant distance, a cubic regression spline basis was used, while for the off-axis angles a two-dimensional tensor product based smooth was used, to incorporate interactions between the horizontal and vertical angles. The largest detection distance for a click was at around 6.5 km, and the model was fitted truncating all data for which distance was larger than 8 km. The maximum degrees of freedom for the smooth of slant distance were set to 4 as, *a priori*, a relatively smooth decreasing function of distance was expected. The models were implemented using the library mgcv in R version 2.8.0 (R Development Core Team, 2008).

Given the fitted detection function model, there are, in general, two approaches for using this model to predict average probability of detection for the primary dataset. If the values of the covariates used in the model are also known for each detection in the primary dataset, then detection probability can be predicted conditional on these values and an average taken. However, if the covariate values for each detected click in the primary dataset are not known, but the *distribution* of the covariate values is known, then its effect can be integrated out of the model, giving the correct average detection probability. This is the approach taken in conventional distance sampling with the distance covariate [Eq. (6)].

In the current study, the values of slant distance and orientation were not known for any of the clicks in the primary dataset. Hence, a simulation approach was used to approximate the distribution of these covariates, based on the motion data from the DTags in the auxiliary dataset. The underlying assumptions were as follows: (i) the paths recorded in the 21 DTag dives from the auxiliary dataset are representative of the paths during the primary survey period, and (ii) the initial position of dives is random with respect to hydrophone position during the primary survey period. To implement the simulation, the following procedure was repeated many times:

(1) Generate the hypothetical location of the first click in a deep dive (x_i, y_i); locations were generated randomly in a buffer area of 8 km around AUTEC.
(2) Sample a dive from the 21 DTag dives available and select the data corresponding to the position of the whale in space (x, y, z location, pitch, heading and roll) at each click event.
(3) Relocate that dive such that the first click is at (x_i, y_i).
(4) For each click event, obtain the slant distance, *hoa* and *voa* with respect to all hydrophones closer than a distance of 8 km to the whale position.

In practice, instead of sampling each click event, only every 100th click was sampled as this was found to increase the speed of convergence of the algorithm to the true distribution of the covariate values. With this refinement, the distributions stabilized after 1000 simulated dives.

Given the above samples from the distribution of cova-

riates, the detection function model fitted earlier was used to estimate detection probability for each click, and the mean of this was taken as an estimate of average detection probability, $\hat{P}$.

To estimate the variance in $\hat{P}$ a nonparametric bootstrap procedure was implemented (e.g., Efron and Tibshirani, 1993). In each of 250 bootstrap replicates, a sample of dives was taken with replacement from the set of 13 dives from which the original detection function had been estimated, the GAM was refitted to this bootstrap resample, and the new fitted model used to produce a new estimate of $\hat{P}$. The empirical variance obtained over these bootstrap estimates is an estimate of the original variance of $\hat{P}$ (see pp. 82–83 of Buckland *et al.*, 2001, for further details in a similar context). This approach assumes that the dives are statistically independent.

2. Estimating cue production rate

The cue production rate (mean number of clicks per whale per second) was estimated using the DTag data. To avoid bias, the periods used for estimating the cue rate were restricted to complete deep-dive cycles (from the start of a deep dive to the start of the next), for which both sound and positional data were available, in a total of 21 deep dives from five whales.

The mean click rate was estimated as a weighted mean of click rate per deep-dive cycle, weighted by deep-dive cycle time length. The variance of this weighted mean was estimated using Cochran's approximation as recommended by Gatz and Smith (1995).

3. Estimating false positive proportion

To estimate the proportion of false positives in the main dataset, a sample of 30 systematically spaced 10 min periods was selected for analysis. For each of the 10 min sample periods, all hydrophones with detections were manually examined using a binary spectrogram display program (Moretti *et al.*, 2006). One of the sample periods was corrupted due to a system test being conducted by range technicians, and was removed from further analysis.

Experienced analysts classified the clicks and assigned click counts into one of the following seven categories: Md, delphinid, sperm whale, boat, pinger, other (which could include other beaked whale species), or "mixed" (i.e., when, in a given minute, Md clicks and other sound sources were not unambiguously distinguishable). Hence the false positive proportion might be between two extreme values, depending on whether all or none of the mixed detections is considered Md clicks. For the density estimates presented here both are used, leading to upper and lower bounds for density, with the best estimate of density likely lying somewhere in between these. The false positive proportion was estimated as the weighted mean of the proportion of false positives by sample period, with total number of clicks detected by sample period as weights. The variance of this weighted mean was estimated using the same procedure as described for the mean click rate variance.

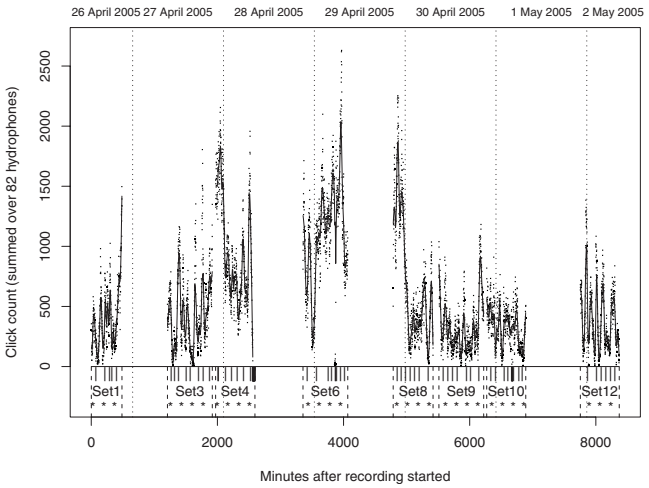


FIG. 2. Click counts per minute, summed over the 82 recording hydrophones, for the 6 day period of the primary dataset. Time is indexed as minutes since recording started. For operational reasons the data were divided into sets, and some sets (2, 5, 7 and 11) were not used. A standard lowess smooth of click counts over time is shown for the sets used. The small black vertical dashes ("|") are scattered minutes within the 8 sets used which were faulty and hence removed from the data. The sample periods used for the estimation of the false positive proportion are represented by " *". The limits of each day and set are represented by dotted and dashed lines, respectively.

C. Case study results

After removing minutes with problems, click counts were available for 4961 min in the primary dataset (Fig. 2).

Pooled over the 82 recording hydrophones, a total of 2 940 521 sounds classified as being beaked whale clicks were detected. Considering hydrophones as independent spatial replicates, the standard deviation of the total number of sounds was 161 373, resulting in a *CV* of 5.5%.

The estimated click rate was 0.407 (clicks/second), with a standard error of 0.040, resulting in a *CV* of 9.8%.

Overall, of the 160 302 sounds detected and considered to be beaked whale clicks during the sample periods used for the false positive proportion estimation, only 78 450 (corresponding to slightly under 50%) were unambiguously identified as Md clicks. The complement of the proportion of false positives, $1-\hat{c}$, was estimated to be 0.549 ($se=0.011$, $CV=1.99\%$) or 0.489 ($se=0.011$, $CV=2.29\%$), depending on whether the mixed clicks are considered to be all or none from Md.

The estimated detection function is shown in Fig. 3. The maximum distance at which a click was detected was 6504 m. The estimated mean detection probability for clicks produced within 8 km of the hydrophones (clicks outside this buffer are assumed to have 0 detection probability) was 0.032, with an associated *CV* of 15.9%.

Using Eq. (3) the estimated Md density at AUTEC over the recording period was 25.3 or 22.5 animals per 1000 km^2 (with respectively 19.5% and 19.6% *CV*'s), depending on the $\hat{c}$ used. Assuming a log-normal distribution for the density estimate, the respective 95% confidence intervals for density are 17.3–36.9 and 15.4–32.9. Note that the contribution of n, $1-\hat{c}$, $\hat{P}$, and $\hat{r}$ to the overall variance estimate on density is

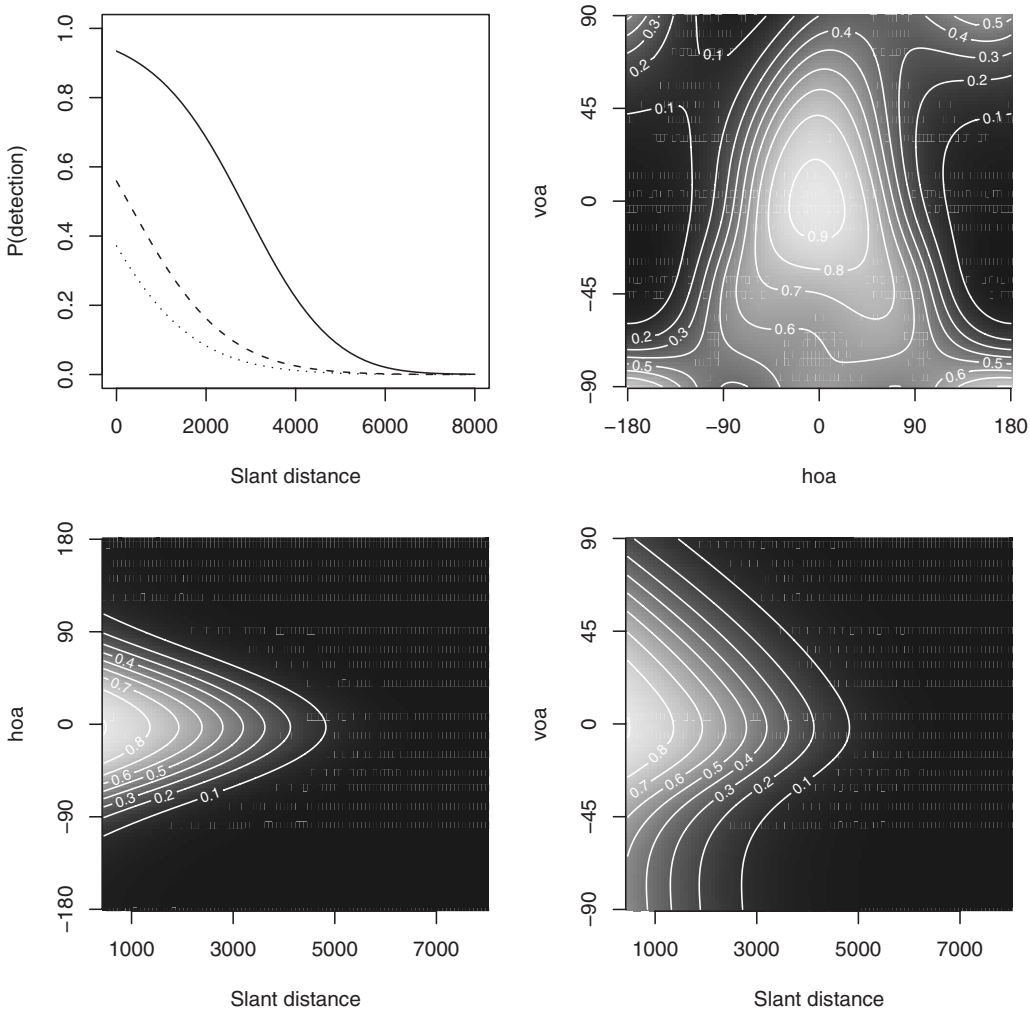


FIG. 3. The estimated detection function. Partial plots (on the response scale) of the fitted smooths for a binomial GAM model with slant distance and a 2D smooth of *hoa* and *voa*. For the top left plot, the off-axis angles are fixed at 0, 45, and 90° (respectively the solid, dashed, and dotted lines). Remaining plots are two-dimensional representations of the smooths, where black and white represent respectively an estimated probability of detection of 0 and 1. Distance (top right panel) and angle not shown (bottom panels) are fixed respectively at 0 m and 0 °.

respectively, 8, 1, 66, and 25%.

IV. DISCUSSION

A. Assessment of case study

1. Comparison with previous work on beaked whales

The case study provided estimates of Md density at AUTEC over a short time period of 6 days. Although true density is not known with certainty, the estimates obtained are consistent with other sources. The vast majority of confirmed beaked whale sightings in the Northern Bahamas are for Md (Claridge, 2006), but the presence of Cuvier's beaked whale (*Ziphius cavirostris*) and Gervais's Beaked Whale (*Mesoplodon europaeus*) have recently been noted at the range.[1] Clicks potentially from other species were treated as false positives, and hence the density estimate presented is valid for Md only.

Using a method based on isolating and counting groups of diving whales over a fixed time period, Moretti *et al.* (2006) obtained estimates of beaked whales at AUTEC using the same primary dataset as analyzed in this paper. Despite their method having some *ad hoc* components, requiring strong assumptions about animal movement and an independent estimate of mean group size, the reported density estimates of 22.0 or 30.0 (depending on the estimate for mean group size used) Md per 1000 km^2 are consistent with the estimates obtained here. (Following a more recent, unpublished, revision of the their estimation process, the area covered by the survey was revised, leading to point estimates of 25.4 or 34.7 Md per 1000 km^2.)

The estimated detection function model (Fig. 3) was consistent with what was expected *a priori*: a decreasing function of distance and a decreasing function of the off-axis angles. This reflects the intuitive notions that the further away and off-axis a Md click is, the harder it is to detect. This also represents further evidence for the high directionality of beaked whale clicks, as has been demonstrated for Cuvier's beaked whale (Zimmer *et al.*, 2005).

The detection of a click seems to be slightly more likely if "below" the whale (negative *voa*) than "above" the whale (positive *voa*): an observation that deserves further investigation in characterizing the Md beam pattern.

At first sight, our results could seem to contradict those from Zimmer *et al.* (2008), which used the passive sonar

equation to model the maximum detection range of Cuvier's beaked whale clicks using a shallow hydrophone in the Mediterranean sea, and predicted a maximum detection range of 4 km. The pertinent parameters utilized were source level of 200 dB re 1 μPa at 1 m, 40 kHz center frequency and processing bandwidth, depth of 0 m, spectral noise level of 30 dB re 1 μPa^2 Hz^{-1}, absorption coefficient of 9.5 dB km^{-1}, and receiver threshold of 14 dB. The deeper hydrophones at AUTEC reduce transmission loss and masking noise, and the processing bandwidth at AUTEC was smaller. In addition, there is justification for assuming a higher maximum source level, since the measurements on which the 200 dB level is based is thought to be an underestimation (Zimmer *et al.*, 2005), and since source levels are known to vary considerably within and between individuals (Zimmer *et al.*, 2008). Johnson *et al.* (2006) reported center and −10 dB frequencies for Md measured at close range using acoustic tags of 38.3 and 26.6 kHz, and hence source levels of 210 and 200 dB re 1 μPa at 1 m at these two frequencies are plausible. Other pertinent parameters are 24 kHz processing bandwidth, depth of 1300 m (approximately mid-way between whale and hydrophone), spectral noise level of 26 dB re 1 μPa^2 Hz^{-1}, absorption coefficients of 8.9 and 5.1 dB km^{-1} at the two frequencies, and receiver threshold of 14 dB. These parameters provide maximum detection ranges of 5.7 km at 38.3 kHz and 7.6 km at 26.6 kHz. The largest observed detection range of a tagged animal in the case study was 6504 m; the above simplistic modeling exercise (e.g., assuming only spherical spreading) suggests that while the high frequency components of the click will be lost at such ranges, sufficient energy remains at lower frequencies to trigger the detector used. Supporting this, the mean received center frequency for clicks detected at 6.5 km range or greater was 29.7 kHz.

2. Assessment of detection probability estimation

Average detection probability was estimated in the case study using a sample of tagged whales for which it could be determined whether cues produced by the whale were detected on surrounding AUTEC hydrophones. However, the simple proportion of these cues detected could not be used directly to estimate average detection probability during the primary survey period for two reasons. First, the sample size was small (four whales over 13 dives), so it is unlikely that the distribution of potentially important covariates in the sample, such as distance from hydrophone, was representative of that over the 6 days of the primary survey. Second, the auxiliary (tag) data were collected at a different time period, when detection conditions were likely different.

The general solution to this problem is to construct a model that includes all the variables that affect detectability and for which the distribution of that variable differs between the auxiliary and the primary datasets. Model selection techniques could be used to select among a suite of candidate variables. Average detectability is then estimated for the correct values of these variables at the time of the primary survey. This approach requires that data are available on the variables affecting detectability both at the time of the auxiliary and primary surveys.

In the case study, because of the small sample size of tags, it was felt that variables indexing the whale's orientation and position relative to the hydrophones may differ between auxiliary and primary data. Both orientation and position were thought *a priori* to have an important influence on detectability. Since orientation and position were not known for cues received in the primary survey, a simulation approach was used to estimate the distribution of these variables during the primary survey. The simulation assumed that a whale's initial position was random with respect to hydrophone horizontal distance (because in the simulation whales were initially located at random within the AUTEC range), and that the dive behavior of the 21 dives from five whales for which data were available in the auxiliary dataset reflected the behavior of whales during the primary survey. The assumption that whale location is random relative to hydrophone seems largely reasonable, given that there were a large number of hydrophones throughout the study area. However, one may expect slightly fewer whales than predicted at larger horizontal distances, because some of the edge hydrophones cover areas that are too shallow for Md to dive in. Nevertheless, the overall influence of these "edge effects" is likely small.

There is no evidence against the assumption that the diving behavior recorded on the 21 dives is representative of that for whales in the primary survey, although this is currently untestable. Baird *et al.* (2008) reported no evidence of a tag-on effect, very similar dive times, and absence of diurnal patterns in dive cycles for Md (with tag durations of up to 72 h).

There are three possibly important variables affecting the detection function that were not considered in the case study. The first is potential differences between hydrophones, since it is possible (but not likely) that the distribution of sensitivities of hydrophones used in the auxiliary dataset is not representative of the 82 recording during the primary survey. The second relates to vertical directivity in the hydrophones, as raised by a reviewer. However, a preliminary analysis of the data revealed that including the vertical angle in the model produced inconsistent results. This is likely because most of the clicks within detection range are made within a narrow band of possible vertical angles with respect to the hydrophones, so any effect of vertical directivity was too negligible to estimate. Even if a strong effect existed, the results of the model reported here would have integrated it implicitly, provided the distribution of vertical angles in the DTag data was representative of the distribution in the 6 day dataset. The third, and potentially more important, is differences in ambient noise conditions, since tags can only be applied during calm weather. For example, mean wind speed, recorded at a location close to the AUTEC range, was 6.1 kn (range 1.1–13.8) during the 13 deep dives used for estimating the detection function, but 12.4 kn (range 2.5–20.8) during the primary survey period. Higher wind speed could reduce detectability, and hence the current estimate of average detection probability could be an overestimate. However the extent of the effect given the depth of the hydrophones remains to be determined. One potential approach would be to include variables such as wind speed, or ambient noise mea-

sured at the hydrophone, in the empirical detection function modeling. However, predicting detection probability for the primary survey period would then mean extrapolating significantly outside the range of the data since maximum wind speed was significantly higher in the primary survey period. An alternative would be to analytically model the effect of increased noise on the detection algorithm, but this is unlikely to be as robust as an empirical approach. A more attractive alternative is to take sound samples from the hydrophones during the primary survey period when no clicks were present and under a range of known wind speeds, use these to contaminate the hydrophone sound files used in the auxiliary study, and thereby empirically study the effect of increased noise on detectability. These approaches will be the subject of a future study.

The detection function was estimated considering only direct path detections. If multi-path detections were included in the click counts in the primary data set, then detection probability could have been underestimated and density correspondingly overestimated. While this is unlikely to have occurred, it might be worth to be investigated further.

It would have been better to have the primary dataset and the auxiliary dataset from the same time period. However, this was not possible because there were not readily available sound recordings for the entire range for the days in which there were DTags. Additionally, placement of DTags corresponded to events considerably spread in time and therefore the mean density estimated over these time periods would be less meaningful.

The method used to estimate probability of detection in the auxiliary dataset assumes that there are no false associations between clicks produced by the whale carrying the tag and clicks detected at the hydrophones. Conceivably, especially if there were other animals around the tagged whale, it is possible that some clicks produced by other animals were wrongly associated with detected clicks on the hydrophone. If this happened, one would overestimate the probability of detecting a click, leading to an underestimation of density. However, this false association seems unlikely: of all clicks detected at more than 4 km, 94.8% were at both *hoa* and *voa* of less than 30°; if these were false associations, one would expect that these correspond to a random sample of off-axis angles, rather than to mostly on-axis clicks. Another check would be to compare the received sound intensity at different frequencies within the 24–48 kHz range for the long-range detections to those predicted from a model of sound production and propagation.

3. Assessment of cue production rate estimation

The auxiliary dataset was also used in estimating click production rate, and hence the assumption is required that average click production rate during the time of the auxiliary survey is equal to that in the primary survey. Information reported in Baird *et al.* (2008) and Tyack *et al.* (2006) about the constancy of deep-dive cycles was fundamental to be confident that a single dive rate was adequate for both day and night periods.

Whenever the cue-count approach proposed here is used, there will be the requirement for an unbiased estimate of average cue production rate to convert cue density into animal density. The application of these methods therefore requires further fundamental research in this area, such as evaluation of cue rates and changes in cue rate as a function of time of day (e.g., Baumgartner and Fratantoni, 2008), year (e.g., Whitehead and Weilgart, 1990), or any other covariates (e.g., group size, season, bottom depth, geographic location, etc). Douglas *et al.* (2005) report that differences in worldwide vocal production (for sperm whales) means that estimates of abundance that are based on cue rates need to be based on data from the population of interest, rather than from another population or some global average.

Note that it is not sufficient to have estimates of the click rate (or inter-click interval) of diving whales from only during vocal parts of their dives—a quantity that is relatively easily obtained from hydrophone data. This will grossly overestimate the required cue production rate, since it ignores the silent time spent between dives. For animals that can spend long periods of time silent, long-term tags or (if feasible) focal follows from the surface are the only options, unless individuals can be tracked through individually recognizable vocalizations.

If cue rate is density dependent, these methods become harder to implement, and it becomes more important to estimate cue rate during the primary survey. If cue rate depends on local group size, then it is important to also estimate cue rate during the primary survey. Failing that, it is still possible to estimate mean cue rate provided one has estimates of the relationship between cue rate and group size, and also the distribution of group sizes during the primary survey.

4. Assessment of false positive proportion estimation

In studies where a manual analysis of the data provides a reliable "gold standard," false positive proportion is the easiest component to estimate. The variance of this component is also easily controllable by taking additional random samples for manual analysis until it is down to an acceptable level. In the case study, the contribution to the overall variance (around 1%) from this component was very small. This does not account for variance due to the specific human operator, and if this is believed to be a considerable source of variation one might consider having different operators repeating the task to quantify it. Here such a procedure was not deemed necessary.

Despite the method being general, it must be noted that values estimated under a given scenario are valid only under that scenario: the number of false positives is dependent on the number of sound sources producing false positives. As an example, everything else being the same, in an area where dolphins (a major source of false positives for the FFT detector used) were much more abundant than at AUTEC, one would expect a much larger proportion of false positives than what was found here.

About half of the sounds recorded as beaked whale clicks were estimated to be false positives in the case study. Such a high proportion was unexpected, although a fair proportion of false positives was expected *a priori*. This emphasizes the importance of accounting for false positives as part of density estimation. A second important lesson is that a

high false positive proportion is not necessarily a barrier to reliable density estimation: it is less important what the characteristics of the detector/classifier are than being able to accurately determine them. This is in contrast to other applications of cetacean detectors—for example, as part of a harm mitigation system—where a large number of false positives would lead to mitigation measures being triggered far more often than required.

The inability to unambiguously classify a proportion of the detected clicks led to increased uncertainty in the reported density estimates, with estimates changing by about 10% depending on how mixed clicks are accounted for.

B. Alternative approaches

1. Alternative methods for estimating P in a cue-counting framework

In the case study, the detection function was estimated based on the auxiliary tagging data. However, sound propagation modeling approaches like those used by Zimmer *et al.* (2008) or Stafford *et al.* (2007) might be alternatively used to obtain the detection function. Alternatively, one could implement broadcasting trials, by mimicking the animal's sound, to empirically estimate the detection function. This would naturally require a comprehensive understanding of the sound production and sound characteristics for the species under study. Ideally, multiple approaches would be used as a way to test the robustness of the results. In all of these cases, the detection function will have to be assumed to be representative of the detection function during the sampling period.

The advantages of using conventional distance sampling methods as described in Buckland *et al.* (2001) are as follows: (i) that the detection function is based on the distances to the detected cues, hence there is no need for additional data from tagging studies, and (ii) the detection function is necessarily valid for the conditions under which the data were collected.

For beaked whales at AUTEC such an approach would be hard to implement, because a given click is often not detected on more than one or two hydrophones, but in a setting where acoustic localization of cues is more feasible such an approach is recommended. Alternative methods of obtaining distances, such as using received sound levels or multi-path detections, may prove useful. Nonetheless, under such a setting there are two fundamental assumptions to estimate the detection function which were not required for the case study: (i) the probability of detecting a cue at distance zero must be 1 (or at least known) and (ii) the distribution of distances of all cues, detected or not, pooled across hydrophones, must be triangular (see Buckland *et al.*, 2001). There is an additional shortcoming for the use of conventional distance sampling methods: the distances required are those to the actual cues (i.e., excluding false positives); alternatively the detection function must be the same for true and false positives. The former seems hard to achieve, and the latter might be unreasonable under some scenarios.

As was described previously, had tags been applied to a large random sample of animals, then P could simply be estimated from the proportion of detected clicks, therefore avoiding the complex modeling exercise presented here. For the data at hand, such an estimate would be 0.036 (CV of 0.155), which is similar to the results from the GAM approach. Nevertheless, the simpler approach cannot be recommended in general for such a small number of dives because the assumptions underlying it are much more restrictive. Rather than having to be representative of the spread of the covariates included in a detection function model, the proportion of clicks detected by dive would have to be representative of the average P.

2. Alternative frameworks for estimating density from fixed passive acoustics

The framework presented here focused on estimating density from an analysis of cues, but other approaches can be envisaged based on detecting animals or groups of animals, and the potential for further research in this area is enormous. If animals can be located based on their vocalizations then the use of an approach similar to the "snapshot" point transect method (Buckland, 2006) could be envisaged. The method could be applied directly if animals could be located using single hydrophone data (e.g., Tiemann *et al.*, 2004); if multiple hydrophones were required for localization then additional development would be needed to account for the dependencies between hydrophones necessary to generate a distance. A major advantage of such an approach would be that it does not require an estimate of cue rate, and so is potentially feasible without any auxiliary data. Similar approaches could be used if groups could be isolated and localized, but additional information on mean group size would be required.

A disadvantage of snapshot type methods is that they require the counts to be discretized into intervals separated in time, so as to introduce statistical independence between samples. Alternative methods that explicitly account for the continuous nature of data acquisition but require information about animal movement exist in the historical literature (e.g., Skellam, 1958); however these need further development to account for issues such as imperfect detection.

In situations, such as for Md at AUTEC, where groups of diving animals can be isolated acoustically, even if not accurately localized, then there is the potential to use a cue counting approach based on counting group dive starts rather than individual clicks. At AUTEC, the probability of detecting a group dive start is 1, so there is no need to estimate the component P in Eq. (3). Similarly, false positives are relatively simple to identify. The cue rate required is the rate of dive initiations, and this could be obtained from time-depth-recording tags, or by tracking focal groups. Hence, such an approach seems potentially feasible and is the subject of ongoing research.

If detections of cues, animals, or groups on multiple hydrophones can be associated, then alternative density estimation methods may be possible that are based on mark-recapture, rather than distance sampling methods. Especially noteworthy are spatially explicit mark-recapture methods (Efford *et al.*, 2008). Mark-recapture methods have the advantage, over conventional distance sampling methods, of

not having to assume that the true distribution of animals with respect to samplers is known (Sec. II B), which is unrealistic if a small number of hydrophones is used. However, mark-recapture methods have their own assumptions, and it is not easy *a priori* to decide which will be more adequate in general.

V. CONCLUSIONS

The framework that has been laid out here is applicable to any study where the key components can be estimated: average detection probability, cue rate, and false positive proportion. Although the case study involved an array of 82 hydrophones over a relatively restricted area and time period, the same methods can be used on a far smaller number of hydrophones (with some loss of precision, obviously), or hydrophones distributed over a much larger area and sampling a longer time period.

Passive acoustic methods are potentially most useful for species that are poorly covered by visual methods but that produce loud, distinctive, predictable vocalizations. Towed acoustic (and visual) methods tend to be better where the goal is wide spatial coverage but narrow temporal coverage; fixed acoustic methods tend to be better where density estimates over long time periods are required. However, the optimal modality depends on many factors, including the study goals, target species, and current technology (the last of which is evolving rapidly). Optimal survey design is an important future research field, in terms of (i) optimal modality for a given situation, (ii) the trade-off between efforts on primary and auxiliary surveys (if required), and (iii) the number and placement of sensors.

Although the focus of this paper has been on estimating cetacean abundance using underwater acoustic data, there is no conceptual extra step required for implementing it with sound recording devices in terrestrial environments. Potential examples include forest surveys of songbirds, territorial monkey species such as gibbons, and elephants (e.g., Payne *et al.*, 2003). Fixed passive acoustics methods are potentially invaluable for monitoring density over time at multiple sites without requiring a large amount of human resources, and for this reason will probably become common practice in the future.

Regarding monitoring of cetaceans, to date most studies have used acoustic data as an index of abundance over time or space (e.g., Stafford *et al.*, 2007). This requires strong (and largely untestable) assumptions that the factors relating number of calls detected to density remain constant (see Anderson, 2001, 2003, for a discussion of these issues). Here, such approaches are improved upon by allowing estimation of these factors, leading to inferences about density.

Mellinger *et al.* (2007b) urged the need for the development of statistical methods for estimating cetacean populations acoustically. This paper shows that the foundations for such methods are laid, and hopes to contribute to their widespread use.

ACKNOWLEDGMENTS

Steve Martin and Dave Moretti provided extensive comments which greatly improved the paper. The authors would like to thank other members of the DECAF project for comment and discussions which are reflected throughout the manuscript: Ron Morrissey, Dave Mellinger, David Borchers, and Catriona Harris. Mark Johnson provided help in processing the DTag data. Danielle Harris, Steve Buckland, Walter Zimmer, Jay Barlow and an anonymous reviewer provided useful comments that improved the clarity and readability of the manuscript. Tagging was performed under US National Marine Fisheries Service research Permit Nos. 981-1578-02 and 981-1707-00 to P.L.T. and with the approval of the Woods Hole Oceanographic Institution Animal Care and Use Committee. This work was funded by two partners under the National Oceanographic Partnership Program: the Ocean Acoustics Program of the US National Marine Fisheries Service, Office of Protected Resources, and the International Association of Oil and Gas Producers Joint Industry Programme on Exploration and Production Sound and Marine Life.

[1]This information was provided by Diane E. Claridge, from the Bahamas Marine Mammal Research Organisation.

Anderson, D. R. (**2001**). "The need to get the basics right in wildlife field studies," Wild. Soc. Bull. **29**, 1294–1297.

Anderson, D. R. (**2003**). "Response to Engeman: Index values rarely constitute reliable information," Wild. Soc. Bull. **31**, 288–291.

Au, W. W. L., Mobley, J., Burgess, W. C., Lammers, M. O., and Nachtigall, P. E. (**2000**). "Seasonal and diurnal trends of chorusing humpback whales wintering in waters off western Maui," Marine Mammal Sci. **16**, 530–544.

Aubauer, R., Lammers, M. O., and Au, W. W. L. (**2000**). "One-hydrophone method of estimating distance and depth of phonating dolphins in shallow water," J. Acoust. Soc. Am. **107**, 2744–2749.

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

Barlow, J., and Gisiner, R. (**2006**). "Mitigating, monitoring and assessing the effects of anthropogenic sound on beaked whales," J. Cetacean Res. Manage. **7**, 239–249.

Barlow, J., and Taylor, B. (**2005**). "Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey," Marine Mammal Sci. **21**, 429–445.

Barlow, J., Ferguson, M. C., Perrin, W. F., Ballance, L., Gerrodette, T., Joyce, G., Macleod, C. D., Mullin, K., Palka, D. L., and Waring, G. (**2006**). "Abundance and densities of beaked and bottlenose whales (family Ziphiidae)," J. Cetacean Res. Manage. **7**, 263–270.

Baumgartner, M. F., and Fratantoni, D. M. (**2008**). "Diel periodicity in both sei whale vocalization rates and the vertical migration of their copepod prey observed from ocean gliders," Limnol. Oceanogr. **53**, 2197–2209.

Buckland, S. T. (**2006**). "Point transect surveys for songbirds: Robust methodologies," Auk **123**, 345–345.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (**2001**). *Introduction to Distance Sampling: Estimating Abundance of Biological Populations* (Oxford University Press, Oxford).

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D., and Thomas, L. (**2004**). *Advanced Distance Sampling* (Oxford University Press, Oxford).

Charif, R. A., Clapham, P. J., and Clark, C. W. (**2001**). "Acoustic detections of singing humpback whales in deep waters off the British Isles," Marine Mammal Sci. **17**, 751–768.

Claridge, D. E. (**2006**). "Fine-scale distribution and habitat selection of beaked whales," MS thesis, University of Aberdeen, Scotland.

DiMarzio, N., Moretti, D., Ward, J., Morrissey, R., Jarvis, S., Izzi, A. M., Johnson, M., Tyack, P., and Hansen, A. (**2008**). "Passive acoustic measurement of dive vocal behavior and group size of Blainville's beaked

whale (*Mesoplodon densirostris*) in the Tongue of the Ocean (TOTO)," Can. Acoust. **36**, 166–173.
Douglas, L. A., Dawson, S. M., and Jaquet, N. (**2005**). "Click rates and silences of sperm whales at Kaikoura, New Zealand," J. Acoust. Soc. Am. **118**, 523–529.
Efford, M. G., Borchers, D. L., and Byrom, A. E. (**2008**). "Density estimation by spatially explicit capture-recapture: Likelihood-based methods," in *Modeling Demographic Processes in Marked Populations*, edited by D. L. Thomson, E. G. Cooch, and M. J. Conroy (Springer, New York), pp. 255–269.
Efron, B., and Tibshirani, R. (**1993**). *An Introduction to the Bootstrap* (Chapman and Hall, London/CRC, New York).
Freitag, L. E., and Tyack, P. L. (**1993**). "Passive acoustic localization of the Atlantic bottlenose dolphin using whistles and echolocation clicks," J. Acoust. Soc. Am. **93**, 2197–2205.
Gatz, D. F., and Smith, L. (**1995**). "The standard error of a weighted mean concentration—I. Bootstrapping vs other methods," Atmos. Environ. **29**, 1185–1193.
Hastie, G. D., Swift, R. J., Gordon, J. C., Slesser, G., and Turrell, W. R. (**2003**). "Sperm whale distribution and seasonal density in the Faroe Shetland Channel," J. Cetacean Res. Manage. **5**, 247–252.
Johnson, M. P., and Tyack, P. L. (**2003**). "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," IEEE J. Ocean. Eng. **28**, 3–12.
Johnson, M., Madsen, P. T., Zimmer, W. M. X., de Soto, N. A., and Tyack, P. L. (**2004**). "Beaked whales echolocate on prey," Proc. R. Soc. London, Ser. B **271**, S383–S386.
Johnson, M., Madsen, P. T., Zimmer, W. M. X., de Soto, N. A., and Tyack, P. L. (**2006**). "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation," J. Exp. Biol. **209**, 5038–5050.
Lewis, T., Gillespie, D., Lacey, C., Matthews, J., Danbolt, M., Leaper, R., McLanaghan, R., and Moscrop, A. (**2007**). "Sperm whale abundance estimates from acoustic surveys of the Ionian Sea and Straits of Sicily in 2003," J. Mar. Biol. Assoc. U.K. **87**, 353–357.
McDonald, M. A., and Fox, C. G. (**1999**). "Passive acoustic methods applied to fin whale population density estimation," J. Acoust. Soc. Am. **105**, 2643–2651.
Mellinger, D. K., Nieukirk, S. L., Matsumoto, H., Heimlich, S. L., Dziak, R. P., Haxel, J., Fowler, M., Meinig, C., and Miller, H. V. (**2007a**). "Seasonal occurrence of North Atlantic right whale (*Eubalaena glacialis*) vocalizations at two sites on the Scotian Shelf," Marine Mammal Sci. **23**, 856–867.
Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P., and Matsumoto, H. (**2007b**). "An overview of fixed passive acoustic observation methods for cetaceans," Oceanogr. **20**, 36–45.
Moore, S. E., Stafford, K. M., Mellinger, D. K., and Hildebrand, J. A. (**2006**). "Listening for large whales in the offshore waters of Alaska," BioScience **56**, 49–55.
Moretti, D., DiMarzio, N., Morrissey, R., Ward, J., and Jarvis, S. (**2006**). "Estimating the density of Blainville's beaked whale (*Mesoplodon densirostris*) in the Tongue of the Ocean (TOTO) using passive acoustics," in Proceedings of the Oceans'06 MTS/IEEE-Boston, Boston, MA.
Morrissey, R. P., Ward, J., DiMarzio, N., Jarvis, S., and Moretti, D. J. (**2006**). "Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean," Appl. Acoust. **67**, 1091–1105.
Payne, K. B., Thompson, M., and Kramer, L. (**2003**). "Elephant calling patterns as indicators of group size and composition: The basis for an acoustic monitoring system," Afr. J. Ecol. **41**, 99–107.
Powell, L. A. (**2007**). "Approximating variance of demographic parameters using the delta method: A reference for avian biologists," Condor **109**, 949–954.
R Development Core Team (**2008**). "R: A language and environment for statistical computing," URL: http://www.R-project.org (Last viewed March 2009).
Seber, G. A. F. (**1982**). *The Estimation of Animal Abundance*, 2nd ed. (Griffin, London).
Skellam, J. G. (**1958**). "The mathematical foundations underlying the use of line transects in animal ecology," Biometrics **14**, 385–400.
Stafford, K., Mellinger, D., Moore, S., and Fox, C. (**2007**). "Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999–2002," J. Acoust. Soc. Am. **12**, 3378–3390.
Thomas, L., Laake, J. L., Strindberg, S., Marques, F. F. C., Buckland, S. T., Borchers, D. L., Anderson, D. R., Burnham, K. P., Hedley, S. L., Pollard, J. H., Bishop, J. R. B., and Marques, T. A. (**2006**). "Distance 5.0," Research Unit for Wildlife Population Assessment, University of St. Andrews, UK. http://www.ruwpa.st-and.ac.uk/distance/(Last viewed March, 2009).
Tiemann, C. O., Porter, M. B., and Frazer, L. N. (**2004**). "Localization of marine mammals near Hawaii using an acoustic propagation model," J. Acoust. Soc. Am. **115**, 2834–2843.
Tyack, P. L., Johnson, M., Soto, N. A., Sturlese, A., and Madsen, P. T. (**2006**). "Extreme diving behaviour of beaked whales," J. Exp. Biol. **209**, 4238–4253.
Ward, J., Morrissey, R., Moretti, D., DiMarzio, N., Jarvis, S., Johnson, M., Tyack, P., and White, C. (**2008**). "Passive acoustic detection and localization of *Mesoplodon densirostris* (Blainville's beaked whale) vocalizations using distributed bottom-mounted hydrophones in conjunction with a digital tag (DTag) recording," Can. Acoust. **36**, 60–66.
Whitehead, H., and Weilgart, L. (**1990**). "Click rates from sperm whales," J. Acoust. Soc. Am. **87**, 1798–1806.
Wiggins, S. M., McDonald, M. A., Munger, L. A., Hildebrand, J. A., and Moore, S. E. (**2004**). "Waveguide propagation allows range estimates for North Pacific right whales in the Bering Sea," Can. Acoust. **32**, 146–154.
Wood, S. N. (**2006**). *Generalized Additive Models: An Introduction With R* (CRC, Boca Raton, FL/Chapman and Hall, London).
Zimmer, W. M. X., Johnson, M. P., Madsen, P. T., and Tyack, P. L. (**2005**). "Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)," J. Acoust. Soc. Am. **117**, 3919–3927.
Zimmer, W. M., Harwood, J., Tyack, P. L., Johnson, M. P., and Madsen, P. T. (**2008**). "Passive acoustic detection of deep diving beaked whales," J. Acoust. Soc. Am. **124**, 2823–2832.

Fundamental constraints on the performance of broadband ultrasonic matching structures and absorbers

O. Acher
CEA DAM, Le Ripault, BP 16, 37260 Monts, France

J. M. L. Bernard
DPTA, CEA DAM, 91297 Arpajon, France

P. Maréchal
LOMC, FRE CNRS 3102, FANO FR CNRS 3110, Place Robert Schuman, 76610 Le Havre, France

A. Bardaine
CEA DAM, Le Ripault, BP 16, 37260 Monts, France

F. Levassort[a)]
Université François-Rabelais, UMR Imagerie & Cerveau, INSERM U930, CNRS ERL 3106, 10 Boulevard Tonnellé, BP 3223, 37032 Tours Cedex 1, France

(Received 21 July 2008; revised 2 December 2008; accepted 22 January 2009)

Recent fundamental results concerning the ultimate performance of electromagnetic absorbers were adapted and extrapolated to the field of sound waves. It was possible to deduce some appropriate figures of merit indicating whether a particular structure was close to the best possible matching properties. These figures of merit had simple expressions and were easy to compute in practical cases. Numerical examples illustrated that conventional state-of-the-art matching structures had an overall efficiency of approximately 50% of the fundamental limit. However, if the bandwidth at −6 dB was retained as a benchmark, the achieved bandwidth would be, at most, 12% of the fundamental limit associated with the same mass for the matching structure. Consequently, both encouragement for future improvements and accurate estimates of the surface mass required to obtain certain desired broadband properties could be provided. The results presented here can be used to investigate the broadband sound absorption and to benchmark passive and active noise control systems. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3081529]

PACS number(s): 43.38.Fx, 43.20.El, 43.20.Hq [AJZ] Pages: 1995–2005

I. INTRODUCTION

Ultrasound techniques are widely employed for medical or non-destructive testing applications, and their use frequently requires layers or multilayers with specific characteristics. It is often desirable to optimize the transmission coefficient between the transducer and the propagation medium. For this, quarter-wavelength matching layers constituted of stacked materials are to a large extent used to improve the coupling between piezoelectric transducers and the propagation medium.[1] While it is relatively simple to fabricate a narrowband adapter, the achievement of a broadband coupling efficiency is often preferred. Efficient absorbers are also required in ultrasonic systems in order to suppress unwanted reflection from the walls of a water tank. The optimization of matching structures and absorbers is related to the physics of waves in layered media, which is a very general problem. The concepts of reflection and transmission, impedance, and wave velocity are common for electromagnetic, acoustic, and electronic applications. Electromagnetic waves as well as sound waves can be described using a common matrix formalism.[1–3] All these fields have experienced a beneficial cross-fertilization over the years. For instance, the perfectly matched layer (PML) approach developed by Bérenger[4] for electromagnetic waves is now used in acoustics.[5,6] As another example, the concept of active absorbers has been proposed in the acoustic community[7] as well as in the field of electromagnetics.[8] Metamaterials and cloaking devices are other examples of recent concepts that present shared interests in these communities.[9–12]

Recently, Rozanov[13] authored a major advance concerning the understanding of the ultimate performance of multilayered electromagnetic absorbers. He proved that the broadband performance of an absorber could be no better than a quantity related to its overall thickness. This property is connected with the causality principle, and is consequently very general. The obtained result was comparable to those established by Bayard–Bode and extended by Fano[14] with regard to broadband-matching of electronic networks.

The purpose of the present paper was to investigate the transposition of this result to acoustics, particularly to the case of ultrasonic matching structures and absorbers. The study was developed on the basis of one-dimensional models at normal incidence. Typical examples of matching layers for piezoelectric transducers and ultrasonic absorbing walls are presented in Sec. II, for which the broadband reflection and

[a)]Author to whom correspondence should be addressed. Electronic mail: levassort@univ-tours.fr

TABLE I. Physical parameters used to calculate the reflection and transmission properties of a PZT/adaptation layer(s)/water structure. (Composition: W-epoxy: epoxy resin loaded with tungsten particles [volume fraction of tungsten powder (%)]; Z: acoustical impedance; ρ: density; v: longitudinal wave velocity; α: attenuation coefficient; d: thickness).

Medium	Composition	Z (MRa)	ρ (kg m^{-3})	v (m s^{-1})	α (dB mm^{-1} MHz^{-1})	d (mm)
Ambient	PZT	33.9	7700	4400	0	∞
Final medium	Water	1.5	1000	1480	0	∞
Single layer $H1$	W-epoxy (19%)	7.1	4670	1520	0.48	0.380
Bilayer $H2$	W-epoxy (36%)	12.0	7740	1550	0.47	0.387
	W-epoxy (4.5%)	4.0	2050	1950	0.40	0.487
Single layer $I1$	W-epoxy (4.5%)	4.0	2050	1950	0.40	0.487
Bilayer $I2$	W-epoxy (3.5%)	3.7	1870	2000	0.39	0.500
	Epoxy	2.8	1240	2260	0.32	0.565

transmission coefficients were obtained through numerical approaches. Section III describes the derivation of general theoretical results concerning the reflection and transmission coefficients of multilayered ultrasonic structures. It is shown that these results become simplified in the cases of a matching structure, as well as for an absorbing structure. The theoretical predictions have been compared to numerical results, and this comparison is discussed in Sec. IV.

II. DESIGN OF MATCHING LAYERS AND ABSORBERS

In the following, numerical examples are used to introduce the general problems treated in the theoretical part. Moreover, they also provide benchmarks in order to confront the theoretical results.

A. Matching structure

In view of being used for ultrasonic transducer applications, piezoelectric materials such as PZT are widely employed. Let us consider a semi-infinite PZT material separated by a plane from a semi-infinite water tank, with a sound wave arising from the PZT side under normal incidence on the interface. Because of the large acoustic impedance mismatch between the two materials (Table I), there was a significant reflection at the interface, and the transmission was limited. In this paper, the complex reflection coefficient associated with the amplitude of the reflected pressure wave is denoted r and the complex transmission coefficient (associated with the amplitude of the transmitted pressure wave) is denoted t (see the Appendix). Moreover, the intensity reflection coefficient R is employed. It can be expressed in decibels and is related to r (see the Appendix) according to

$$R = 10\log_{10}|r|^2 = \frac{10}{\log(10)}\ln|r|^2. \tag{1a}$$

A similar relation is used to define the coefficient T from t by

$$T = 10\log_{10}|t|^2. \tag{1b}$$

A very general case consists in a matching stack inserted between an ambient medium, indexed a, and a final medium, indexed b, with a sound wave originating from medium a under normal incidence [Fig. 1(a)]. In order to improve the energy transfer, a common approach is to insert a quarter-wavelength matching layer with an acoustic impedance ranging between the two materials or a multilayer [Figs. 1(a) and 1(b)] with appropriate thicknesses. For transducer applications, this gives rise to an increase in their sensitivity.

In the so-called "harmonic adaptation" approach, the acoustic impedance Z of a quarter-wavelength matching layer is indexed H and given by

$$Z_{H(N,n)} = (Z_a)^P (Z_b)^{1-P} \quad \text{with} \quad P = \frac{N+1-n}{N+1}, \tag{2}$$

where N is the total number of quarter-wavelength matching layers, and n is the considered number of these layers, with $1 \le n \le N$. Moreover, Z_a is the impedance of the PZT layer

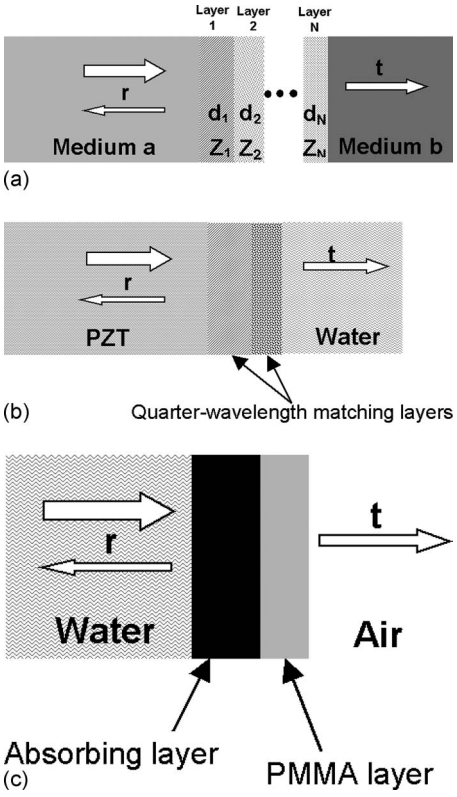


FIG. 1. (a) A sketch of a multilayered (N) structure between incident (a) and propagation (b) media; (b) a sketch of a PZT/water structure with an adaptation layers; (c) a sketch of a water tank coated with an absorbing layer.

(ambient), and Z_b that of the water (final medium), according to the notations in Fig. 1(a).

More particularly, in the case of a single quarter-wavelength matching layer, the acoustic impedance becomes

$$Z_{H(1,1)} = (Z_a Z_b)^{1/2}. \tag{3a}$$

A thicker structure with two-quarter-wavelength matching layers is expected to provide a better bandwidth. In this approach, the impedance of the two layers should be

$$Z_{H(2,1)} = Z_a^{2/3} Z_b^{1/3} \quad \text{and} \quad Z_{H(2,2)} = Z_a^{1/3} Z_b^{2/3}. \tag{3b}$$

In the so-called "impulse regime" approach, a broadband condition is formalized through the binomial transformer formula.[15] The acoustic impedance of the considered quarter-wavelength matching layer is indexed I. One can demonstrate that it can be written as follows:

$$Z_{I(N,n)} = (Z_a)^P (Z_b)^{1-P} \quad \text{with} \quad P = \frac{2^{N+1-n} - 1}{2^{N+1} - 1}. \tag{4}$$

More particularly, a single quarter-wavelength matching layer should verify

$$Z_{I(1,1)} = Z_a^{1/3} Z_b^{2/3}. \tag{5a}$$

A two-quarter-wavelength matching layer structure in this approach corresponds to

$$Z_{I(2,1)} = Z_a^{3/7} Z_b^{4/7} \quad \text{with} \quad Z_{H(2,2)} = Z_a^{1/7} Z_b^{6/7}. \tag{5b}$$

These acoustic impedance values can be obtained in composite materials where a tungsten powder is dispersed in an epoxy resin.[16] The ultrasonic characteristics of the various materials are summarized in Table I.

The modeling of the studied structure was developed on the basis of a one-dimensional assumption and took into account multiple reflections between the interfaces. Thus, a Debye series expansion based on the Fresnel reflection and transmission coefficients led to unconditionally stable formulas. By using the notation proposed by Conoir,[17] a recursive approach could be implemented.[18] The R and T coefficients of these structures are shown in Fig. 2. A significant reduction in the unwanted reflection coefficient at the central frequency 1 MHz was observed in all cases, along with an improvement of the transmission coefficient. It was clearly observed that two-quarter-wavelength matching layers—as opposed to a single quarter-wavelength matching layer—increased the bandwidth, but also rendered the matching structure thicker and heavier.

B. Absorber

Let us consider a semi-infinite water tank separated from the air by a 1-cm thick PMMA planar wall (polymethyl methacrylate polymer or its corresponding commercial Plexiglas™), with an ultrasonic wave arising under normal incidence. Because of the large impedance mismatch between the water and the air, a very large reflection occurred unless an ultrasonic absorber was placed against the PMMA wall [Fig. 1(c)]. Figure 3 presents the intensity reflection coefficient R for an elastomeric absorber that was either 30 or 60 mm thick, with properties reported in Ref. 19. In this example, increasing the absorber thickness improved the low-frequency operation and increased the overall efficiency. The different material characteristics are reported in Table II.

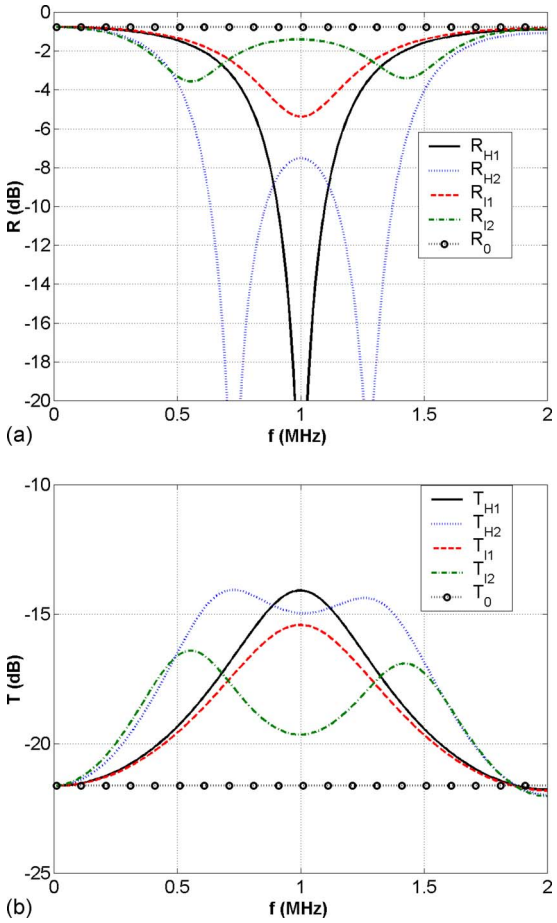


FIG. 2. (Color online) The reflection R and transmission T coefficients associated with the PZT-water interface (circles: R_0 and T_0) and with various matching structures tuned at 1 MHz (solid lines: $H1$; dotted lines: $H2$; dashed lines: $I1$; dashed-dot lines $I2$; see Table I for further details).

III. THEORETICAL APPROACH

The main focus was placed on the very general problem of broadband adaptation. It became clear that perfect match-

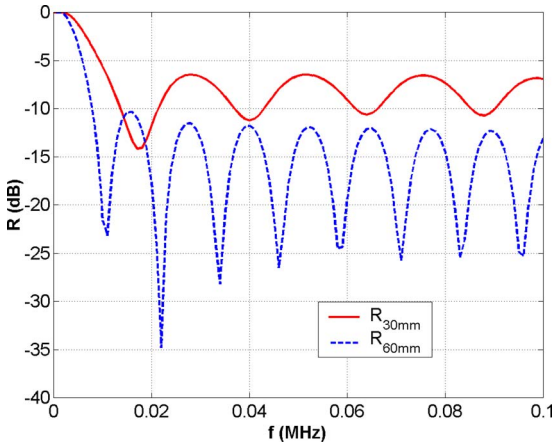


FIG. 3. (Color online) The reflection coefficient R at a water/absorber/PMMA/air interface, for absorber thicknesses of 30-mm (solid line) and 60-mm (dashed line).

TABLE II. Physical parameters used to calculate the reflection and transmission properties of a water/absorber/tank/air structure. (Z: acoustical impedance; ρ: density; v: longitudinal wave velocity; α: attenuation coefficient; d: thickness).

Medium	Composition	Z (MRa)	ρ (kg m^{-3})	v (m s^{-1})	α (dB mm^{-1} MHz^{-1})	d (mm)
Ambient	Water	1.5	1000	1480	0	∞
Final medium	Air	4×10^{-4}	1.2	330	0	∞
Tank wall	PMMA	3.5	1270	2750 @ 2 MHz	$0.13+0.94\times(\mathrm{F/MHz})$	10
Absorber	Loaded elastomer	1.8	1150	1670 @ 2 MHz	$0.74\times(\mathrm{F/MHz})^{1.38}$	30 60

ing in order for no reflection to occur was possible at a single frequency, as evidenced in Fig. 2 for the case $H1$. However, would it be possible to design a sandwich of adaptation layers operating over a significant bandwidth, so that no reflection would take place over the entire frequency operating range? If not, are there design rules that indicate how thick, or how heavy, the adaptation sandwich should be in order to attain some given figure of merit in terms of adaptation? Similar questions can be raised, concerning the transmission coefficient. A simple mathematical solution to a perfect broadband-matching layer would be to take an adaptation structure that works at a single frequency, such as $H1$, with the requirement that the index varies as $1/f$, for which the impedance is constant with the frequency f. However, such a material would be impossible to manufacture since it does not respect causality.

In the following, the general problem of broadband adaptation is treated with the constraint that the materials in the matching layers respect the causality principle, which limits the variation of their properties with the frequency. This problem of broadband adaptation has been solved for electromagnetic waves in 1998.[13,20,21] The seminal work of Rozanov[13] established the relation between the broadband efficiency of microwave absorbers and the minimum thickness that was necessary in order to obtain this efficiency. It has provided a sound figure of merit for the absorbers and an accurate basis for comparing various absorber types. This approach was successfully extended to broadband antenna substrates, as well as to high impedance planes.[22] It has further been very useful for anticipating the best way to use new material concepts, such as metamaterials.[10] Moreover, it can provide benchmarks for comparing passive and active impedance-matching solutions. These relations belong to a large class of results derived from the causality principle, such as Kramers–Kronig, and they can be viewed as the counterparts of Bayard–Bode and Fano results[14] related to impedance-matching in electronics. However, the derivation provided by Rozanov[13], employed in the present study, differs from the Fano approach.

In accordance with previous work,[13] the aim of the present study was to investigate the integral of the intensity reflection coefficient R over the wavelength spectrum. As an illustration, Fig. 4(a) depicts the reflection coefficient associated with the case $H1$ reported in Table I. It shows not only the dip of the reflection coefficient at the central operation frequencies but also the behavior of this quantity over an extremely broad bandwidth. For very large wavelengths, the reflection coefficients tend toward that at a PZT/water interface, denominated $R_0=10\log_{10}|r_0|^2$ (see the Appendix). Contrarily, at small wavelengths, interference effects lead to numerous peaks. The aim of the derivation presented below was to determine the integral of the hatched area. In the case presented here, the integral was clearly negative. Due to the fact that it is much more common to describe the reflection coefficient as a function of the frequency, rather than as a function of the wavelength, the derivation was carried out

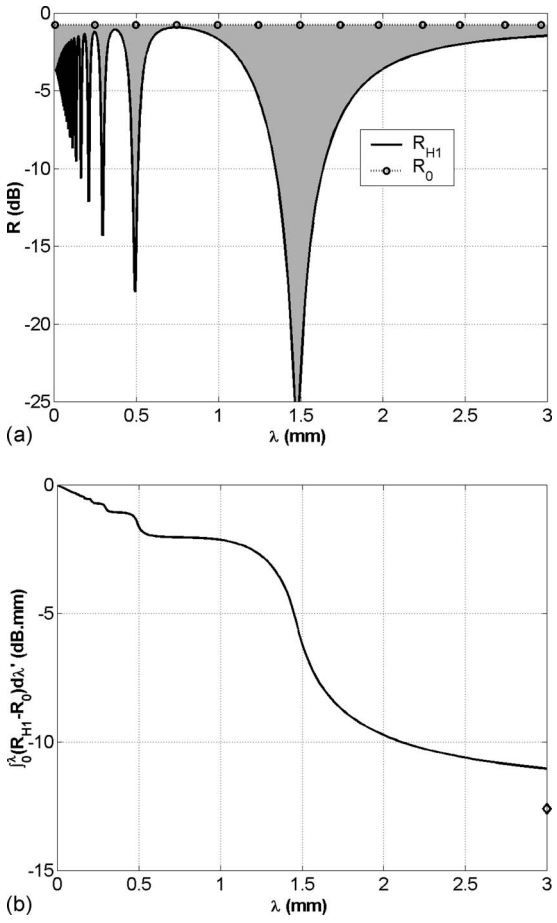


FIG. 4. (a) The reflection coefficient in PZT in the presence of an adaptation layer $H1$ between PZT and water (solid line R_{H1}); the wavelength is calculated in the water; the straight dashed line with circles represents the asymptotic value R_0 of the reflection coefficient. (b) The integral of the reflection coefficient over the wavelength in water, corresponding to the gray area in Fig. 4(a). The symbol (diamond) on the right axis represents the asymptotic value of the integral.

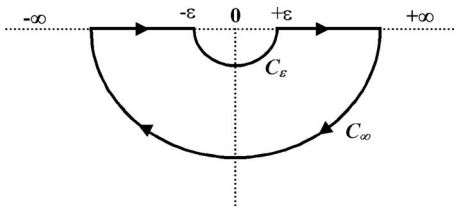


FIG. 5. A sketch of the integration contour in the complex plane.

with all literal expressions as a function of the frequency. Moreover, with the exception of Fig. 4, all results are displayed as functions of the frequency.

A. General case

The pressure reflection coefficient $r(f)$ as a function of the complex frequency f is analytic over the lower half plane,[5] and this property is due to causality.[13,14] Our main interest was directed toward the properties of $\log(r(f))$; however, this function was not analytic due to the possible presence of zeros of the pressure reflection coefficients. The zeros of r in the lower half plane should be denominated f_q, where $1 \leq q \leq Q$, and Q is the number of poles. Rozanov[13] introduced an ancillary function $r'(f)$ with two attractive properties: for real frequencies $|r(f)|=|r'(f)|$ and $\log(r'(f))$ is analytical over the lower half plane. $r'(f)$ is defined by

$$r'(f) = r(f)\prod_{q=1}^{q=N}\frac{f-f_q^*}{f-f_q} = r(f)\frac{(f-f_1^*)}{(f-f_1)}\frac{(f-f_2^*)}{(f-f_2)}\cdots\frac{(f-f_Q^*)}{(f-f_Q)}, \tag{6}$$

where the asterisk stands for the complex conjugation.

At zero frequency, the pressure reflection coefficient tends toward a real limit r_0. The quantity of interest is the integral of $\log(r(f)/r_0)$ over the wavelength range λ. Since $\lambda=c/f$, where c is the celerity of the waves, it corresponds to the integral over f of $\log(r(f)/r_0)/f^2$, within a proportionality factor. Using the aforementioned properties of $r'(f)$, and taking into account that $r'(-f^*)=r'(f^*)$, the quantity of interest is

$$\int_0^\infty \log\left|\frac{r(f)}{r_0}\right|^2\frac{df}{f^2} = \int_{-\infty}^\infty \log\left(\frac{r'(f)}{r_0}\right)\frac{df}{f^2}. \tag{7}$$

The right-hand part of Eq. (7) is the integral of a function that is analytical in the lower half of the complex plane, except in zero. As a consequence, the integral of this quantity over a closed contour sketched in Fig. 5 is zero. The integral over the real axis corresponds to the right member of Eq. (7), which is the quantity to be calculated. The integral over the circle C_∞ tends toward zero when the radius of this circle increases, because of the $1/f^2$ factor in the function to be integrated. The integral over the circle C_ε can be expressed as a function of the derivative of the pressure reflection coefficient at null frequency. It follows that

$$\int_0^\infty \log\left|\frac{r(f)}{r_0}\right|^2\frac{df}{f^2} = \pi\,\mathrm{Im}\left(\frac{d(r(f))}{r_0 df}\right)_{F=0} + 2\pi\,\mathrm{Im}\left(\sum_q \frac{1}{f_q}\right). \tag{8}$$

Since the f_q are in the lower half plane,

$$\mathrm{Im}\left(\sum_q \frac{1}{f_q}\right) \geq 0. \tag{9}$$

This gives

$$\int_0^\infty (R-R_0)\frac{df}{f^2} \geq \frac{10\pi}{\log(10)}\mathrm{Im}\left(\frac{d(r(f))}{r_0 df}\right)_{f=0}. \tag{10}$$

The derivation of the first order development of $r/r_0(f)$ in f around $f=0$ for a multilayer is given in the Appendix. It is expressed as a function of the acoustic impedances Z_a and Z_b of the ambient and final media, of the thicknesses d_n of the various layers, of their acoustic characteristics in the low-frequency limit, the impedance Z_n, and the speed of sound c_n. Alternatively, Z_n can be expressed as a function of the density ρ_n, through $Z_n=\rho_n.c_n$. Consequently,

$$\int_0^\infty (R-R_0)\frac{df}{f^2} \geq -\frac{40\pi^2}{\log(10)}\frac{Z_a}{Z_a^2-Z_b^2}(m-Z_b^2h) = -K_{RR}, \tag{11}$$

where

$$m = \sum_n d_n\frac{Z_n}{c_n} = \sum_n d_n\rho_n, \tag{12}$$

$$h = \sum_n \frac{d_n}{Z_n c_n} = \sum_n \frac{d_n}{\rho_n c_n^2}, \tag{13}$$

and $40\pi^2/\log(10)\approx 171$.

Here, m is the mass per unit of surface of the matching layers.

It should be noted that the left-hand side of Eq. (11) is generally negative, where an increasing negativity corresponds to an enhanced antireflection efficiency. The right-hand side of Eq. (11) is also a negative number, which is indeed a prediction of the performance bounds of the multilayer structure in terms of antireflection efficiency.

A similar approach can be worked out for the transmission coefficient T

$$\int_0^\infty (T-T_0)\frac{df}{f^2} \geq -\frac{20\pi^2}{\log(10)}\frac{(m+Z_aZ_bh)}{Z_a+Z_b}. \tag{14}$$

B. Case of a transducer matching structure

In the case of a matching structure attached to a transducer, a number of simplifications can be made. In particular, $Z_a \gg Z_b$ and $Z_b^2h \ll m$ as calculated in Table III for harmonic adaptation [in other cases Eq. (11) must be used]. As a consequence, Eq. (11) becomes

TABLE III. Calculated parameters (h, Z_b^2h, m, and Z_aZ_bh) for various adaptation layers and absorbing structures.

Configuration	h ($m^2\ s^2\ kg^{-1}$)	Z_b^2h ($kg\ m^{-2}$)	m ($kg\ m^{-2}$)	Z_aZ_bh ($kg\ m^{-2}$)
		Adaptation structure		
$H1$	3.5×10^{-14}	0.08	1.8	
$H2$	8.3×10^{-14}	0.18	3.9	
$I1$	6.2×10^{-14}	0.14	1.0	
$I2$	15.6×10^{-14}	0.35	1.6	
		Absorber structure		
30-mm	1×10^{-11}	1.9×10^{-6}	47	0.0065
60-mm	2×10^{-11}	3.7×10^{-6}	82	0.013

$$\int_0^\infty (R-R_0)\frac{df}{f^2} \geq -\frac{40\pi^2}{\ln(10)}\frac{m}{Z_a}. \tag{15}$$

Equation (15) is a remarkably simple expression. It proves that whatever the arrangement of layers in a matching structure, a certain minimum weight is required in order to obtain the desired antireflection properties. However, it should be kept in mind that this is a bound, and that it, naturally, does not prevent the building of very inefficient absorbers from high mass walls. In the case where the material parameters m and Z_a are not precisely known, it is useful to notice that the relative uncertainty of the limit is simply the sum of the relative uncertainties on m and Z_a. The relation associated with the transmission is not relevant in this case, since the integral quantity related to the transmission in Eq. (14) is not an adequate figure of merit for high transmission structures.

C. Case of an absorber at a water/air interface

In the case of an absorber at a water/air interface, $Z_a \gg Z_b$ and $Z_aZ_bh \ll 1$ as calculated in Table III for our studied cases. It follows that

$$\int_0^\infty (T-T_0)\frac{df}{f^2} \geq -\frac{20\pi^2}{\ln(10)}\frac{m}{Z_a} = -K_{RT}. \tag{16}$$

It simply reads that the integrated efficiency of an absorber at the water/air interface, in order to absorb waves propagating in the water, is bounded by the mass of the absorber. Rozanov's bound for the transmission coefficient T is half of that for the reflection coefficient R, i.e., $K_{RT}=K_{RR}/2$. Efficiency to minimize the transmitted waves from water to air is described by the left-hand side of Eq. (16).

IV. DISCUSSION

It is useful to compare the theoretical predictions with the various numerical examples that were investigated in the first part of this study. Such a comparison should provide a check of the general predictions on particular examples and introduce a more general discussion.

A. Matching structure

Figure 6(a) illustrates the reflection coefficient R associated with the $H1$ structure at the PZT-water interface (Table I), and Fig. 6(b) displays that the argument of the integral in Eq. (11) rapidly tended toward zero as f increased. As a consequence, it can be seen in Fig. 6(c) that the integral was very close to its asymptotic value when the upper integration frequency was above the third resonance. This statement was consistent with what can be observed in Fig. 4(b), where the partial integral was represented as a function of the wavelength in water: the contribution of wavelengths shorter than the third resonance was found to be negligible. Rozanov's

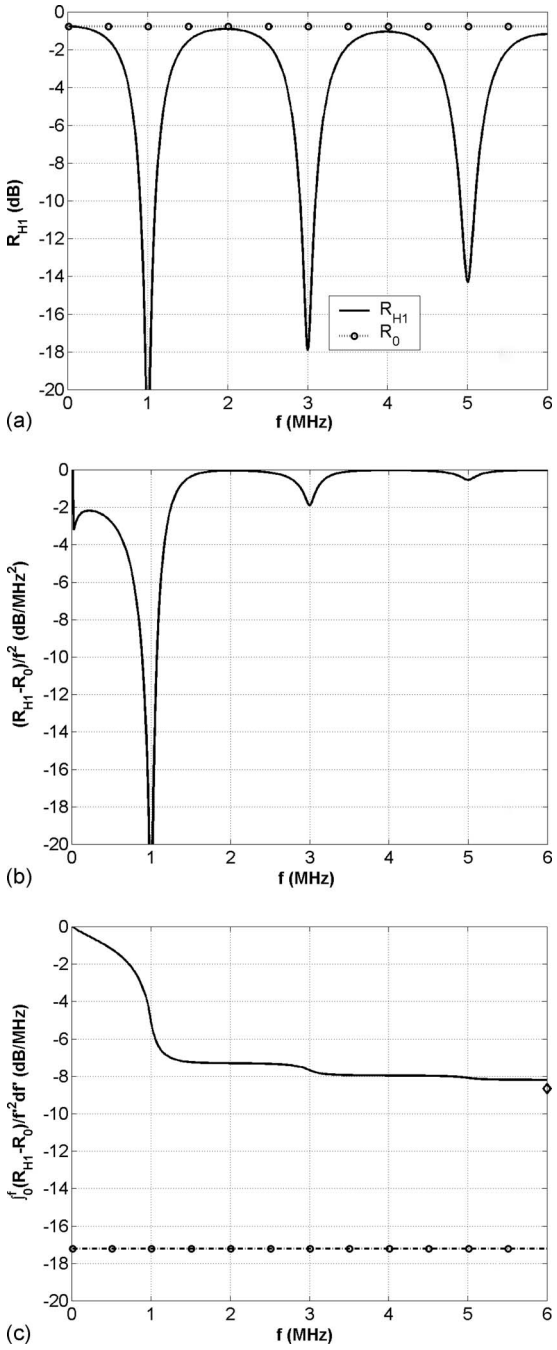


FIG. 6. (a) A reflection coefficient through PZT in the presence of an adaptation layer $H1$ between PZT and water, as a function of the frequency (solid line R_{H1}). The limit R_0 at zero frequency is indicated by the dashed line with circles. (b) The reflection coefficient R_{H1} weighted by $1/f^2$. (c) The corresponding partial integral (solid line) and comparison with the Rozanov limit (dashed line with circles). The symbol (diamond) on the right axis correspond to the summation up to 250 MHz.

TABLE IV. Values of Rozanov bounds K_{RR}, figures of merit η and η', and bandwidth $\Delta f/f$ for the matching structures.

Adaptation structure	K_{RR} (dB MHz^{-1})	$\left\lvert \int_0^\infty (R-R_0)\frac{df}{f^2} \right\rvert$ (dB MHz^{-1})	η (%)	$\Delta f/f$ at -6 dB (%)	η' (%)
$H1$	17	8.65	50	34	10
$H2$	39	19.4	50	87	12
$I1$	9	4.3	49	0	0
$I2$	13	6.3	48	0	0

bound K_{RR} associated with the integral of the reflection coefficient R was calculated using Eq. (15), and is presented in Table IV. It is also displayed in Fig. 6(c), and the prediction of Eq. (15) was clearly fulfilled.

The three other impedance-matching structures described in Table I were also investigated. Their reflection coefficients R are presented in Fig. 7(a), and the integrals as well as Rozanov's bounds are pictured in Fig. 7(b). It can be seen that the integral quantity is in module about a factor 0.5 from Rozanov's bound. Table IV summarizes Rozanov's bounds for the various structures. It is quite straightforward to understand that the increased broadband efficiency of the $H2$ structure was strongly connected with its considerably heavier surface weight of 3.9 kg m^{-2} (Table III), as compared to the much less efficient $I1$ structure with a weight of only 1 kg m^{-2}. These results also suggested that certain possibilities remained in order to increase the antireflection efficiency with the same mass, through a more clever arrangement of materials, and possibly the use of dispersive matter.

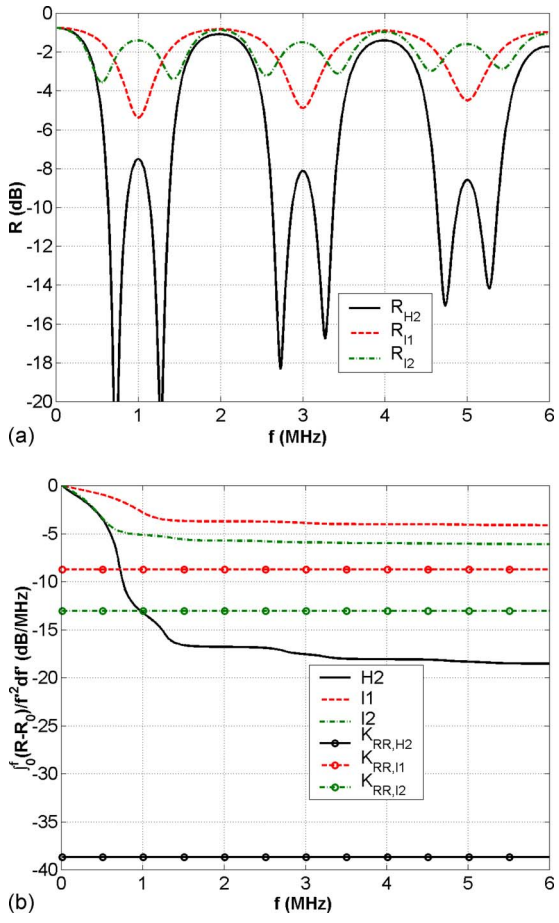


FIG. 7. (Color online) (a) The reflection coefficients through PZT in the presence of various adaptation layers between PZT and water, as functions of the frequency ($H2$: solid line; $I1$: dashed line; $I2$: dashed-dot line). (b) The corresponding partial integrals and a comparison with their respective Rozanov limits K_{RR} (horizontal lines with circles).

B. Absorber

Let us consider the absorbing structure described in Fig. 1(c) and Table II, for which the reflection coefficient R is given in Fig. 3. The integral associated with the reflection coefficient R is presented in Fig. 8, and Table V provides the calculated value of Rozanov's bounds. It was expectedly found that, for each absorber thickness, the integral did not go as far in the negative values as the ultimate bounds according to Eq. (15). Significant room for optimization was expected in the present case, since the integral was in module about 20 times less than Rozanov's bound. The transmission coefficients T for the two absorber thicknesses are given in Fig. 9(a), and their associated integrals can be seen in Fig. 9(b). Each integral was approximately five times less in module than its associated value of K_{RT}.

C. Fundamental and practical uses of Rozanov's relations

It has been established that the wideband efficiency of antireflection layers, either in the matching layer configura-

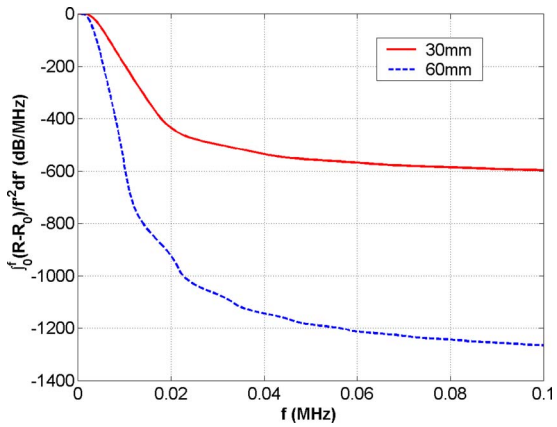


FIG. 8. (Color online) The partial integral associated with the reflection coefficient at a water/absorber/PMMA/air interface, for an absorber thickness of 30-mm (solid line) or 60-mm (dashed line). The reflection coefficient R is displayed in Fig. 3.

TABLE V. Values of Rozanov bounds K_{RR} and K_{RT} for the absorber structures.

Absorber thickness (mm)	K_{RR} (dB MHz^{-1})	K_{RT} (dB MHz^{-1})
30	11 500	5 800
60	20 000	10 000

tion or in the absorbing layer configuration, was bounded by a simple expression K_{RR}. In practical cases, the bound is proportional to the surface mass of the structure, divided by the impedance of the ambient media. This provides straightforward hints for the engineering of ultrasonic matching and absorbing layers. The main point is that it affords a clear figure for the minimum budget required to build the necessary reflection properties. This budget is not calculated in money, but in kg m^{-2}. The second point is that it provides an indicator for the efficiency of the designed structure. The following ratio η is a good candidate for this

$$\eta = \frac{-1}{K_{RR}} \int_0^\infty (R - R_0) \frac{df}{f^2}. \tag{17}$$

η cannot exceed unity as a consequence of Eq. (11). The closer the ratio is to 100%, the closer the solution is to the ultimate limits. Examples have been presented above, where this ratio could range from nearly 50% (the matching structures) down to 5% (with the elastomeric absorber).

In a number of practical cases, the benefits brought by a matching or an absorbing multilayer are summarized as a certain frequency band $[f_1, f_2]$ where the intensity reflection coefficient R has been reduced below a certain level $R_{\min}$ expressed in decibels. Since $R \leq R_{\min}$ in the interval $[f_1, f_2]$, it follows, according to Eq. (11), that

$$0 \geq (R_{\min} - R_0) \frac{f_2 - f_1}{f_1 f_2} \geq \int_0^\infty (R - R_0) \frac{df}{f^2} \geq -K_{RR}. \tag{18}$$

It is then possible to establish

$$(f_2 - f_1)(R_0 - R_{\min}) \leq K_{RR} f_1 f_2. \tag{19}$$

In the case where the bandwidth $\Delta f = f_2 - f_1$ is relatively small compared to the center resonance frequency $f_r = (f_1 + f_2)/2$, it is possible to obtain a simple expression of the maximum bandwidth that can be brought by a certain mass of the matching or absorbing multilayer. Subsequently, Eq. (19) reduces to

$$\Delta f (R_0 - R_{\min}) \leq K_{RR} f_r^2. \tag{20}$$

It is sometime preferred to use the relative bandwidth $\Delta f / f_r$, and the bound for this quantity also has a simple expression

$$\frac{\Delta f}{f_r} \leq K_{RR} \frac{f_r}{R_0 - R_{\min}}. \tag{21}$$

In the same way as for Eq. (17), it is possible to obtain a simple figure of merit associated with a matching or an absorbing multilayer

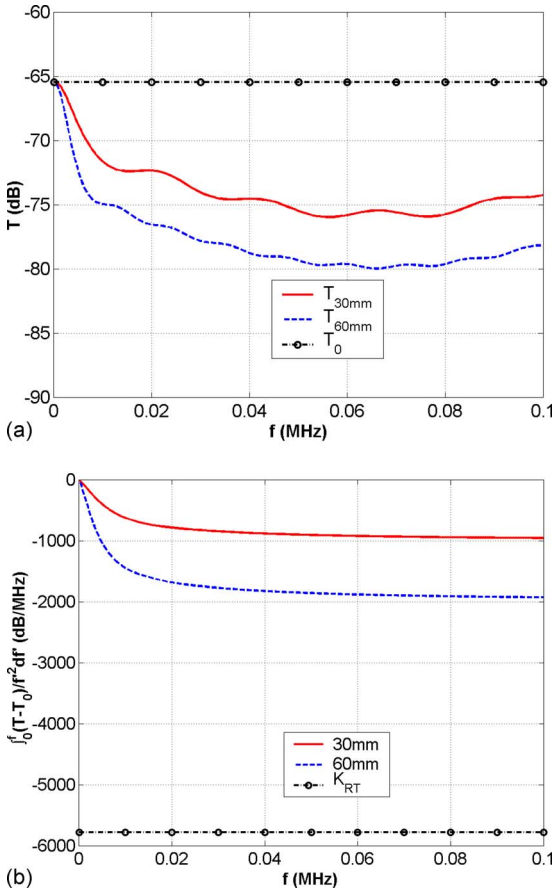


FIG. 9. (Color online) (a) The transmission coefficient at a water/absorber/PMMA/air interface, for an absorber thickness of 30-mm (solid line) or 60-mm (dashed line). (b) The corresponding partial integral associated with the transmission coefficient (30-mm: solid line; 60-mm: dashed line) and the Rozanov limit K_{RT} for the 30-mm case (horizontal dashed line with circles).

$$\eta' = \frac{\Delta f (R_0 - R_{\min})}{K_{RR} f_r^2}. \tag{22}$$

η' is always lower than unity as a consequence of Eq. (21). It is enough to know the bandwidth Δf at $R_{\min}$ (decibels), the center frequency f_r, the mass m of a multilayer, and the characteristics of the initial and final medium to compute the figure of merit η'.

An illustration related to the practical use of Rozanov's bounds in association with bandwidth consideration is provided in Fig. 10. This illustration is associated with the quarter-wavelength matching structure $H2$ (already presented in Figs. 2 and 7, as well as in Tables I, III, and IV). Figure 10 displays the reflection coefficient R close to the matching frequency $f_r = 1$ MHz. It can be seen that the bandwidth at -6 dB of the structure $H2$ is [0.57 MHz, 1.44 MHz]. The gray area is the quantity $\Delta f (R_0 - R_{\min})$, corresponding to the left-hand side of Eqs. (19) and (20). This quantity is bounded by Rozanov's constant times the square of the central frequency, according to Eq. (20). In this example, $\Delta f (R_0 - R_{\min}) = 4.6$ MHz dB, while the bound is 39 MHz dB: the prediction of Eq. (20) is thus fully verified. The gray area is $\eta' = 12\%$ of the ultimate limit, given the mass that has been involved in building the $H2$ structure.

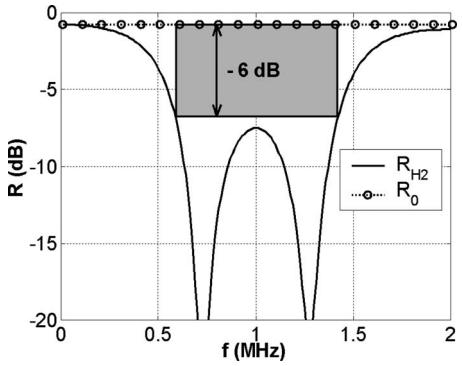


FIG. 10. The reflection coefficient associated with the matching structure $H2$ at the PZT-water interface; the bandwidth at $R_{min}=-6$ dB is also sketched, along with the reflection coefficient R_0 in the absence of a matching layer. The gray area is bounded, according to Eq. (20).

The adimensional figures of merit η and η' are useful clues in order to detect which cases can be significantly improved by further design efforts, and which ones cannot. They may also be useful for comparing passive structures and active sound absorbers,[23] as well as for guiding the optimization process of broadband sound absorbers.[24] In the case of very miniature systems, such as silicon-based microelectromechanical system, the requirements of a significant surface mass may be a limitation in order to obtain good matching properties. In this case, the theoretical results presented here suggest various paths to cope with this limitation. For instance, a very abrupt reflection curve should be synthesized, with very little remaining attenuation outside the working bandwidth, especially at lower frequencies. The high frequency tail gives negligible contribution to the integral, whereas the low-frequency tail, corresponding to large wavelengths, may significantly contribute to the integral, and therefore lead to larger masses, as can be seen in Fig. 6. Table IV indicates that the integral on the left-hand side of Eq. (15) reaches about $\eta=50\%$ of its maximal allowed value for a given mass, whereas the bandwidth at -6 dB is, at most, $\eta'=12\%$ of its maximal allowed value for the same given mass.

From a more fundamental point of view, Rozanov's bound provides a good benchmark for comparing very different approaches for impedance-matching, including passive and active methods. Further, it raises fundamental questions that are not answered by the present work. Is it possible to attain these bounds? If so, what are the material properties required for this task? Is the influence of the frequency dispersion of the material properties important? Is it possible to derive global optimization criteria that include the piezoacoustic transducer and its driving electronics? The existence of a matrix formalism describing the entire system,[25] as well as the existence of similar relations both in electronics[14] and for ultrasounds, suggests that such unified figures of merit may exist. It is believed that the acoustic community can further adapt these results to its specific theoretical and applied problems, including broadband sound absorption and active noise control.[23,24]

D. Common points and differences with the electromagnetic case

Although the formal description is the same, significant differences exist between Rozanov's relations for the electromagnetic and ultrasonic cases. The existence of a reference medium (vacuum) for the celerity of the electromagnetic waves renders it more customary to describe electromagnetic waves through their wavelength. As a consequence, the expression of Rozanov's relations in terms of an integral over the wavelength is probably more suitable in the electromagnetic domain than in the ultrasonic field.

While the expressions of Rozanov's bound are the same when written as a function of the impedances, indices, and thicknesses of the various layers, they become very different when expressed as functions of other physical parameters. In Eqs. (11), (15), and (16), m is the surface mass in the case of acoustics, while it corresponds to the sum of the permeability by thickness product of the different layers for electromagnetic materials. A sum rule related to the surface mass of electromagnetic absorbing materials was derived, but associated with the integral over f, rather than over λ.[26]

It seems much more common to use materials with a strong frequency dispersion of the electromagnetic properties[21] as opposed to materials with a strong frequency dispersion of the acoustic properties. However, the advent of metamaterials in the sonic range may provide new and attractive techniques to carve sharp ultrasonic resonance features that could prove appealing.[10]

The canonical cases are also quite different in the various domains. In electromagnetics, one always considers $Z_b=0$, which is a fully adequate description of conductors, and therefore one obtains $|r_0|=1$, and $R_0=10\log_{10}|r_0|^2=0$ dB. The acoustic impedance-matching structure for an ultrasonic transducer departs from that configuration, and it is consequently necessary to establish the integral relations taking into account a nonzero R_0. In the case of the absorbing wall between water and air, R_0 is close to 0 as in the electromagnetic case. However, the transmission into air may be a concern since air does not dampen the acoustic waves. As a consequence, the relation associated with the transmission coefficient, Eq. (16), makes sense.

V. CONCLUSION

The ultimate performance of matching and absorbing structures used in ultrasonic applications is bounded by the surface mass. This result is very general, since it was derived from the causality principle. The bounds provided practical indications that could be applied to the conception of multilayer structures, and the relations gave rise to certain figures of merit that could come in useful for comparing passive and active absorbers or matching structures. These figures of merits are simple to compute in practical cases, requiring only the knowledge of the bandwidth and of the mass of the matching structure. Nevertheless, certain questions remain unanswered; for instance, whether it is possible

to attain these bounds, and, if so, with which structure. In addition, further adaptations to the case of broadband sound absorption and noise control are expected to lead to fruitful developments.

APPENDIX

Let us consider a plane wave incoming under normal incidence on a multilayer structure as sketched in Fig. 1(a). The pressure P and local displacement speed v between two faces of layer 1 are related by

$$\begin{pmatrix} P(d_1) \\ v(d_1) \end{pmatrix} = \overline{\overline{M}}_1 \begin{pmatrix} P(0) \\ v(0) \end{pmatrix}, \tag{A1}$$

with

$$\overline{\overline{M}}_1(f) = \begin{pmatrix} \cos\left(\frac{2\pi f d_1}{c_1}\right) & -jZ_1 \sin\left(\frac{2\pi f d_1}{c_1}\right) \\ -j/Z_1 \sin\left(\frac{2\pi f d_1}{c_1}\right) & \cos\left(\frac{2\pi f d_1}{c_1}\right) \end{pmatrix}, \tag{A2}$$

where Z_1 and c_1 are the acoustic impedance and sound celerity for the medium 1 at the frequency f, and d_1 is its thickness. Since only the first order development in f is required in Eq. (10), it is convenient to express the first order development of the above matrix as

$$\overline{\overline{M}}_1(f) = \overline{\overline{I}} - j2\pi f \begin{pmatrix} 0 & \frac{Z_1 d_1}{c_1} \\ \frac{d_1}{Z_1 . c_1} & 0 \end{pmatrix} + \overline{\overline{O}}(f^2). \tag{A3}$$

Here, $\overline{\overline{I}}$ is the identity matrix and $\overline{\overline{O}}(f^2)$ designates terms that are in f^2 and higher orders. Z_1 and c_1 are the low-frequency limits of the acoustic impedance and sound celerity. For the whole multilayer, the pressure and velocity fields at the output interface of the multilayer, denoted with the index n, are related to the fields at the first interface, denoted with the index 0, through

$$\begin{pmatrix} P_n \\ v_n \end{pmatrix} = \overline{\overline{M}} \begin{pmatrix} P_0 \\ v_0 \end{pmatrix} \quad \text{with} \quad \overline{\overline{M}} = \prod_{i=1}^{n} \overline{\overline{M_{n-i+1}}} = \overline{\overline{M}}_n \ldots \overline{\overline{M}}_1, \tag{A4}$$

resulting in

$$\overline{\overline{M}}(f) = \overline{\overline{I}} - j2\pi f \begin{pmatrix} 0 & m \\ h & 0 \end{pmatrix} + \overline{\overline{O}}(f^2), \tag{A5}$$

where m is the surface mass defined by

$$m = \sum_{i=1}^{n} \frac{Z_i d_i}{c_i} = \sum_{i=1}^{n} d_i \rho_i \tag{A6}$$

and

$$h = \sum_{i=1}^{n} \frac{d_i}{Z_i c_i}. \tag{A7}$$

The fields at interfaces 0 and n can be expressed as functions of the pressure field P_i associated with the incoming sound wave

$$\begin{pmatrix} P_0 \\ v_0 \end{pmatrix} = \begin{pmatrix} 1-r \\ \frac{1+r}{Z_a} \end{pmatrix} P_i \quad \text{and} \quad \begin{pmatrix} P_n \\ v_n \end{pmatrix} = \begin{pmatrix} t \\ \frac{t}{Z_b} \end{pmatrix} P_i. \tag{A8}$$

Therefore,

$$\begin{pmatrix} t \\ \frac{t}{Z_b} \end{pmatrix} = \overline{\overline{M}} \begin{pmatrix} 1-r \\ \frac{1+r}{Z_a} \end{pmatrix} \tag{A9}$$

and it follows that

$$\frac{r}{r_0} = 1 - j2\pi f \frac{2Z_a}{Z_a^2 - Z_m^2}(m - Z_b^2 h) + O(f^2), \tag{A10}$$

$$\frac{t}{t_0} = 1 - j2\pi f \frac{(m + Z_a Z_b h)}{Z_a + Z_b} + O(f^2), \tag{A11}$$

where

$$r_0 = \frac{Z_a - Z_b}{Z_a + Z_b} \quad \text{and} \quad t_0 = \frac{2Z_b}{Z_a + Z_b} \tag{A12}$$

are, respectively, the reflection and transmission coefficients in the absence of the multilayer.

[1] B. G. Martin, "Theory of the acoustical of symmetric multilayer structures," J. Appl. Phys. **66**, 1931–1934 (1989).
[2] B. A. Auld, *Acoustic Fields and Waves in Solids*, 2nd ed. (R. E. Krieger, Malabar, Florida, 1990).
[3] O. Acher, A. L. Adenot, and F. Duverger, "Fresnel coefficients at an interface with a lamellar composite material," Phys. Rev. B **62**, 13748–13756 (2000).
[4] J. P. Bérenger, "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200 (1994).
[5] X. Yuan, D. Borup, J. W. Wiskin, M. Berggren, R. Eidens, and S. A. Johnson, "Formulation and validation of Berenger's PML absorbing boundary for the FDTD simulation acoustic scattering," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 816–822 (1997).
[6] G. Wojcik, M. Reshef, and D. Loewenthal, "Elastic wave calculations by the Fourier method," Bull. Seismol. Soc. Am. **74**, 875–891 (1984).
[7] H. F. Olson and E. G. May, "Electronic sound absorber," J. Acoust. Soc. Am. **25**, 1130–1136 (1953).
[8] K. N. Rozanov and E. A. Preobrazhenskii, "Synthesis of wideband radar absorbers based on complex media composed from active electric dipoles," J. Commun. Technol. Electron. **50**, 787–793 (2005).
[9] G. Veselago and E. E. Narimanov, "The left hand of brightness: Past, present and future of negative index materials," Nature Mater. **5**, 759–762 (2006).
[10] N. Fang, D. Xi, J. Xu, M. Ambati, W. Sriturayanich, C. Sun, and X. Zhang, "Ultrasonic metamaterials with negative modulus," Nature Mater. **5**, 452–456 (2006).
[11] A. Cho, "Cloak of invisibility unveiled," Science **314**, 403 (2006).
[12] G. W. Milton, M. Briane, and J. R. Willis, "Quasistatic cloaking of two-dimensional polarizable discrete systems," New J. Phys. **8**, 248–267 (2006).
[13] K. N. Rozanov, "Ultimate thickness to bandwidth ratio of radar absorbers," IEEE Trans. Antennas Propag. **48**, 1230–1234 (2000).
[14] R. M. Fano, "Theoretical limitations on the broadband matching of arbitrary impedances," J. Franklin Inst. **249**, 57–83 (1950).
[15] C. S. Desilets, J. D. Fraser, and G. S. Kino, "The design of efficient

broadband piezoelectric transducers," IEEE Trans. Sonics Ultrason. **25**, 115–125 (1978).
[16]T. Nguyen, M. Lethiecq, F. Levassort, and L. Pourcelot, "Experimental verification of the theory of elastic properties using scattering approximations in (0–3) connectivity composite materials," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 640–645 (1996).
[17]J. M. Conoir, "Acoustic scattering from cylinders," Ph.D. Thesis, Pierre and Marie Curie (Paris VI), Paris, France (1987).
[18]P. Maréchal, L. Haumesser, L. P. Tran-Huu-Hue, J. Holc, D. Kuščer, M. Lethiecq, and G. Feuillard, "Modeling of a high frequency ultrasonic transducer using periodic structures," Ultrasonics **48**, 141–149 (2008).
[19]P. He, "Experimental verification of models for determining dispersion from attenuation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 706–714 (1999).
[20]K. N. Rozanov, "Fundamental limitation for thickness to bandwidth ratio of radar absorbers," Proceedings of PIERS 1998: Progress in Electromagnetics Research Symposium (1998), Vol. **1**, p. 229.
[21]K. N. Rozanov and S. N. Starostenko, "Numerical study of bandwidth of radar absorbers," Eur. Phys. J.: Appl. Phys. **8**, 147–151 (1999).
[22]C. R. Brewitt Taylor, "Limitation on the bandwidth of artificial perfect magnetic conductor surfaces," IET Proc. Microwaves, Antennas Propag. **1**, 255–260 (2007).
[23]M. Cuesta, P. Cobo, A. Fernandez, and J. Pfretzschner, "Using a thin actuator as secondary source for hybrid passive/active absorption in an impedance tube," Appl. Acoust. **67**, 15–27 (2006).
[24]Y. C. Chang, L. J. Yeh, and M. C. Chiu, "Optimization of constrained composite absorbers using simulated annealing," Appl. Acoust. **66**, 341–352 (2005).
[25]G. F. Engen, "The six-port reflectometer: An alternative network analyzer," IEEE Trans. Microwave Theory Tech. **25**, 1075–1080 (1977).
[26]O. Acher and A. L. Adenot, "Bounds on the dynamic properties of magnetic materials," Phys. Rev. B **62**, 11324–11327 (2000).

The rotary subwoofer: A controllable infrasound source

Joseph Park[a] and Milton Garcés
Infrasound Laboratory, University of Hawaii, 73-4460 Queen Kaahumanu Highway No. 119, Kailua-Kona, Hawaii 96740

Bruce Thigpen
Eminent Technologies, 225 East Palmer Street, Tallahassce, Florida 32301

(Received 14 June 2008; revised 23 January 2009; accepted 24 January 2009)

The rotary subwoofer is a novel acoustic transducer capable of projecting infrasonic signals at high sound pressure levels. The projector produces higher acoustic particle velocities than conventional transducers which translate into higher radiated sound pressure levels. This paper characterizes measured performance of a rotary subwoofer and presents a model to predict sound pressure levels. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3082115]

PACS number(s): 43.38.Ja, 43.28.Dm, 43.20.Rz, 43.28.Hr [AJZ] Pages: 2006–2012

I. INTRODUCTION

While considerable technological progress has focused on infrasound sensors, signal processing, and propagation effects, fundamental physical barriers have hindered the development of a controllable infrasound source.[1] Classic audio transduction technology such as voice-coil activated diaphragms, piezoelectric resonators, and magneto-resistive sheets lack either the large spatial dimensions or vibratory displacements required to effectively radiate high-intensity infrasound.

A massive horncoupled electropneumatic loudspeaker was developed for the U.S. Army Research Laboratory: the mobile acoustic source (MOAS).[2] The MOAS is frequently referred to as the "mother of all speakers" and was capable of projecting high-intensity sound to frequencies as low as 10 Hz, though it was a large, power intensive, complex system. Use of conventional loudspeaker technology in an array has demonstrated the ability to project acoustic signals down to frequencies of 8 Hz;[3] however, the source levels were modest. Inventors have explored more exotic loudspeaker technology, for example, the "hydrosonic subwoofer" used fluid coupling to impedance match infrasound to enclosure vibrations,[4] while optimization of a driven Helmholtz resonator also demonstrated infrasound generation.[5] However, these devices have peculiar needs such as fluid bladders or require physical changes to tune alternate frequencies. Explosive sources (including air guns) can generate sufficient acoustic pressures; however, they are relatively uncontrollable in amplitude and duration, and are not amenable to continuous-wave sound projection. Explosives also entail a host of environmental and safety issues.

An alternative technology has been recently developed consisting of a baffled fan with dynamically controlled blade pitch[6,7] and is referred to as the rotary speaker. An electric motor rotates the fan hub at a constant frequency Ω while the blade pitch is dynamically controlled by a signal $s(t)$ at frequency ω_s (Fig. 1). In the non-radiating state, blade pitch is zero with minimal air displacement. In the radiating mode the blade pitch is actively modulated according to the input signal and the resultant pulses of air create high amplitude coherent oscillations in a controlled manner.

Classic transduction techniques are limited by poor radiation resistance at small values of ka, where k is the wavenumber and a the characteristic dimension of the radiator. For example, a uniformly oscillating piston in a rigid semi-infinite baffle operating at $ka \ll 1$ has a radiation resistance R_r that is close to zero. Efficient radiation does not occur until higher frequencies are reached $(ka > 4)$ where $R_r \approx 1$. This paper demonstrates that the rotary speaker overcomes this limitation by efficiently projecting relatively high sound pressure levels (SPLs) even though the source is compact $(ka \ll 1)$. We speculate that the rotary woofer creates an effective R_r even though $ka \ll 1$ because it changes the radiation impedance from reactance (mass-loading) dominated to resistance (fluid displacement) dominated by virtue of increased particle velocities.

This paper has three main goals: to introduce the rotary subwoofer as a controllable infrasound generator to the acoustics community, to present measurements and performance characterization of the device, and to present a simple theoretical model for the source. This paper is organized as follows. Section II provides a physical description of the source and its installation; Sec. III details the measurements and their results. Characterization of the speaker enclosure is explored in this section to explain increased amplitudes of farfield radiation. Section IV presents an acoustic model of the source based on an effective particle velocity distribution. Results of the model are compared with measurements in Sec. V, and discussion of the results are provided.

II. SPEAKER DESCRIPTION AND INSTALLATION

The rotary speaker pitch mechanism uses a conventional electromagnetic voice-coil assembly driven by an audio amplifier to pitch the blades in response to the applied signal $s(t)$. The voice-coil longitudinal oscillations are converted to blade axis rotary motion within the hub. The tested configuration consisted of five equally spaced blades, each approxi-

[a]Electronic mail: jpark@isla.hawaii.edu

 0001-4966/2009/125(4)/2006/7/$25.00

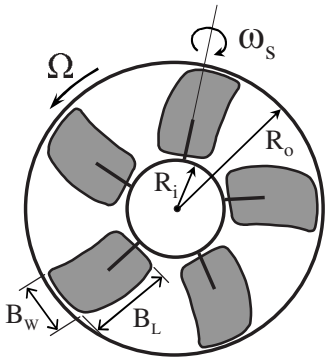


FIG. 1. The rotary subwoofer consists of a hub of radius R_i rotating at a fixed rate of Ω with blades of length B_L and width B_W modulated in pitch by signal $s(t)$ with frequency ω_s.

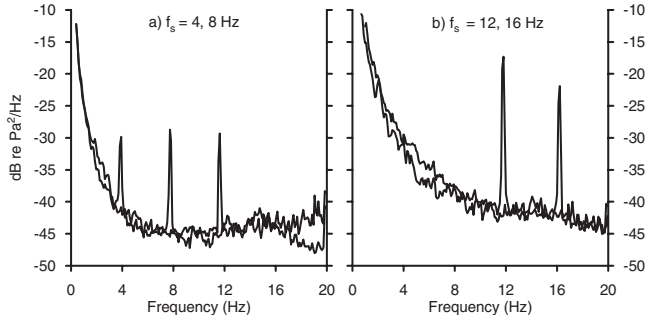


FIG. 3. Acoustic power spectra measured at range of 1820 m with fan rotation frequency of $F=17$ Hz. (a) shows signal frequencies of $f_s=4$ and 8 Hz and (b) $f_s=12$ and 16 Hz.

mately $B_W=10.2$ cm (4 in.) in width by $B_L=15.3$ cm (6 in.) in length. The outer diameter of the hub is approximately 20.4 cm (8 in.) so that the annular region of radiation has inner and outer radii of $R_i=10.16$ cm (4 in.) and $R_o=25.40$ cm (10 in.). The axis of the rotor is approximately 61 cm (24 in.) above the ground.

The speaker is installed at the Infrasound Laboratory with the rotor hub and blades flush with a heavy sound board baffle that is clamped inside an exterior door opening with dimensions 91.4 cm (3 ft) × 213.4 cm (7 ft). The projector radiates into free space outside the building, while the building interior forms the back volume. Figure 2 shows a photograph of the installed source.

The building dimensions are approximately 9.14 m (30 ft) × 18.3 m (60 ft) × 4.3 m (14.2 ft). There are no structural partitions within the building. Construction consists of a bolted steel frame, with W- and S-shape I-beams used as columns and beams. Corrugated steel panels are screwed to the frame to form exterior walls and roof. The result is a building with considerable structural compliance compared to a concrete block design. When the rotary woofer is in operation the exterior wall panels are easily observed to oscillate.

III. MEASUREMENTS

Measurements were conducted with calibrated Chaparral Physics model 2.2A infrasound microphones.[8] Data were recorded and digitized on a RefTek model 130-01 at a sample frequency of 100 Hz. Input signals were generated by an EZ Digital FG-8002 function generator. The blade actuation signal $s(t)$ was a sine wave with constant amplitude and frequency ω_s for any given measurement. Minimum record lengths for each measurement were 120 s. Spectral processing used a 50% data overlap with 1024 point Welch's averaged modified periodogram method and a minimum of 25 averages.

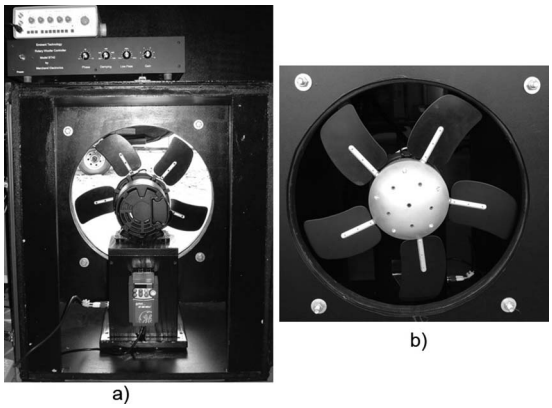


FIG. 2. Photograph of the installed source. (a) view inside the building showing the function generator, amplifier, motor controller, and motor and (b) view from the exterior showing the main rotor hub and blades.

Owing to the small atmospheric absorption of infrasound (less than 0.005 dB/km for frequencies below 16 Hz),[9] we were able to easily detect the source at considerable distance. Figure 3 shows on-axis spectra recorded at a range of 1820 m for signal frequencies of $f_s=4$, 8, 12, 16 Hz with a constant fan rotation of $F=17$ Hz. Figure 3(a) shows the $f_s=4$ and 8 Hz results recorded in the morning with light winds and lower noise levels. The $f_s=4$ Hz signal produced the harmonic at 12 Hz. Figure 3(b) plots the results at $f_s=12$ and 16 Hz during a period of higher winds and higher background noise. During the noisier measurement period the 4 Hz signal would not have been detectable without filtering. Even at a range of 1.8 km, there is impressive signal-to-noise ratio exceeding 20 dB for the 12 and 16 Hz projections.

The acoustic model presented in Sec. IV decomposes the particle velocities at the face of the blades into two components. The first is produced by the rotational velocity of the fan about the main hub at frequency F, the other from the rotational velocity of the blade about its actuation axis at frequency f_s. The velocity component produced from main hub rotation will change in proportion to the frequency F, the radius of the hub, and the outer radius of the blades. The component related to f_s is controlled by the amplitude of the signal $s(t)$ and the rate of blade actuation f_s. As the amplitude of $s(t)$ increases the blade pitch actuation increases driving the blade through larger displacements and higher velocities.

Figures 4 and 5 quantify changes in SPL as a function of the three parameters F, f_s, and $|s(t)|_{\text{rms}}$. To estimate the SPL at a particular signal frequency, a time-series analysis was used. A bandpass filter of 1 Hz was centered on the signal frequency; the computed variance provides an estimate of the signal power and is converted into decibels.

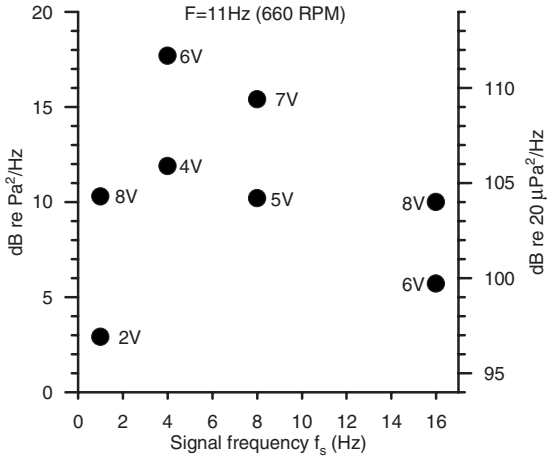


FIG. 4. SPL from measurements at 1 m with $F=11$ Hz versus signal frequency f_s. rms voltage amplitude of the drive signal is denoted next to each measurement.

Figure 4 presents projector SPL measured at 1 m with a constant hub frequency of $F=11$ Hz (660 rpm). Values of root mean square (rms) input signal voltage amplitude ($V_{\text{in}}=|s(t)|_{\text{rms}}$) are denoted next to each estimate. It is clear that at this fan rotation frequency the signal frequencies of $f_s=4$, 8 Hz produce a more efficient response than the lower or higher signal frequencies. It can also be observed that for a specific signal frequency that an increase in signal amplitude (blade pitch actuation) produces an increase in projected SPL, for example, at $f_s=4$ Hz an increase in input signal power from 4 to 6 V_{rms} results in a measured SPL increase of about 5 dB.

Measurements of SPL at 1 m with a hub rotational frequency of $F=17$ Hz (1020 rpm) are shown in Fig. 5. Compared to the $F=11$ Hz results, we find significant increases in SPL. Data represented by solid symbols were recorded at a range of 1 m, while the open symbols were recorded at a range of 1820 m and corrected to 1 m based on spherical spreading loss. Peak levels measured at 1 m were 28.1 dB re Pa^2/Hz (122.2 dB re 20 $\mu\text{Pa}^2/\text{Hz}$).

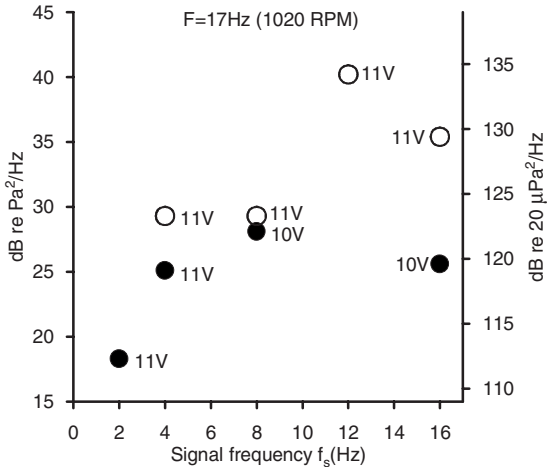


FIG. 5. SPL from measurements at 1 and 1820 m with $F=17$ Hz versus signal frequency f_s. Solid symbols were recorded at a range of 1 m; open symbols were recorded at a range of 1820 m and corrected to 1 m. rms voltage amplitude of the drive signal is denoted next to each measurement.

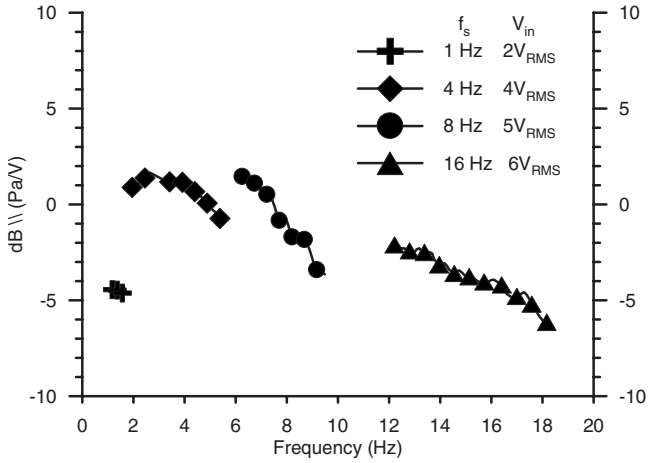


FIG. 6. Composite signal transfer function with values of coherence >0.98.

SPL estimates at 12 and 16 Hz from the range corrected 1820 m measurements are unexpectedly high. Two possible mechanisms for this could be changes in the propagation channel and resonant coupling from the baffle/back volume. It is known that changes in temperature profile and wind conditions can produce a profound impact on the local (<50 km) propagation of infrasound.[10] The measurements at $f_s=12$ and 16 Hz were performed with a southwest (040°) wind of approximately 4 m/s aligned with the direction of the projector axis (032°), while the $f_s=4$ and 8 Hz measurements were collected earlier in the day with little or no wind.

Resonant coupling could arise from forced oscillation of the building close to its natural modes. For a rectangular enclosure of dimensions l_x, l_y, lowest axial mode eigenfrequencies in hertz are $\omega_n=c/\pi l_x$ and $\omega_m=c/\pi l_y$, which evaluate to $f_n=9.4$ Hz ($l_x=18.3$ m) and $f_m=18.8$ Hz ($l_y=9.14$ m). When the enclosure is driven at frequency ω, the axial modes can be described by

$$P_{nm}=\frac{A_{nm}\cos k_n x\cos k_m y}{\sqrt{(1/Q_{nm})^2+(\omega/\omega_{nm}-\omega_{nm}/\omega)^2}}, \tag{1}$$

where A_{nm} is determined by the source and Q_{nm} the quality factor governed by boundary losses.[11] It is assumed that change in eigenfrequency as a result of the building panel oscillation is small since the oscillations are much smaller than the characteristic building dimensions.

The second term in the denominator describes an amplification (α) factor that maximizes P_{nm} when $\omega=\omega_n$. At a signal frequency of $f_s=12$ Hz and building eigenfrequency of $f_n=9.4$ Hz, the value of $\alpha_n=2.03$. With $f_s=16$ Hz, $f_m=18.8$ Hz and $\alpha_m=3.19$. This suggests that signal frequencies of 12 and 16 Hz are close enough to modal frequencies of the building that there is potential for amplification of building panel oscillation at the drive frequencies. It is therefore possible that the building itself is acting as an extended radiator when driven at signal frequencies of 12 and 16 Hz.

Figure 6 plots a composite transfer function of the output acoustic power at a range of 1 m to the input voltage signal power V_{in}. There are four distinct transfer function spectra plotted in Fig. 6, a separate measurement at each of the four drive frequencies $f_s=1$, 4, 8, and 16 Hz, with a

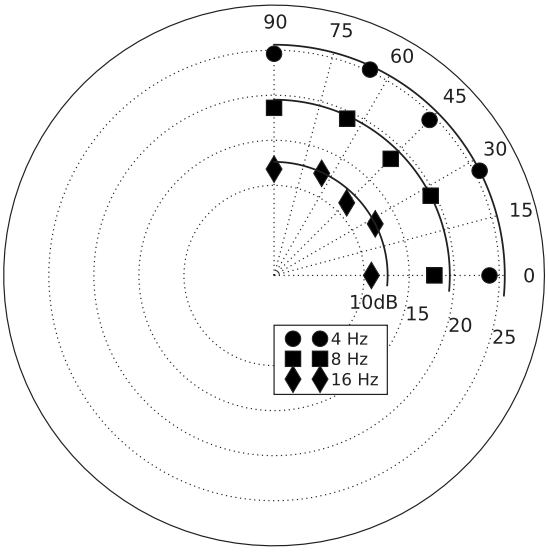


FIG. 7. Angular response measured at a range of 1 m for the drive frequencies f_s=4, 8, 16 Hz. Data are SPL in units of Pa^2/Hz. The axis of the fan is aligned at 0°.

constant fan rotation frequency of $F=11$ Hz. Portions of each spectrum were plotted only if the input/output coherence was greater than 0.98. Interpreting the coherence as a metric of input/output linearity, we can see that as the drive frequency increases the spectral band over which the assumed linear response is valid increases, though in none of the measurements does it cover the entire spectrum. A possible reason is that as the drive frequency decreases the increasing compliance of the diaphragm may decouple the input/output.

Turning attention to the amplitude of the transfer functions, the speaker exhibits maximal power conversion over the 3–8 Hz band at this F. This is consistent with the measurements presented in Fig. 4. A notable result of Fig. 6 are the relatively large amplitudes (>-4 dB), which indicate a high transduction efficiency despite the low value of kR_o.

A five-point angular response of the projector at a range of 1 m is shown in Fig. 7. At the low wavenumbers and small radiator dimensions ($kR_o<0.15$) of the rotary speaker, the radiation field of a baffled simple source should be monopole. The data show a persistent reduction of approximately 2 dB on-axis for all measurements, with similar results at a range of 10 m. Reasons for this departure from the expected omnidirectional radiation are not clear. Evaluation with the source model discussed below indicates that radiation from the wall/baffle is capable of reducing the on-axis pressure.

IV. ACOUSTIC MODEL

The rotary speaker relies on the motion of the hub and blades to generate significant particle velocities, while the dynamically controlled blade pitch creates the fluctuating signal pressure pulses. These two components are similar to those found in cooling/turbo-machinery fans and helicopter rotor blades, both of which have been extensively studied. There are, however, some differences. Fan studies are typically assessed with inclusion of duct or strut interactions; the

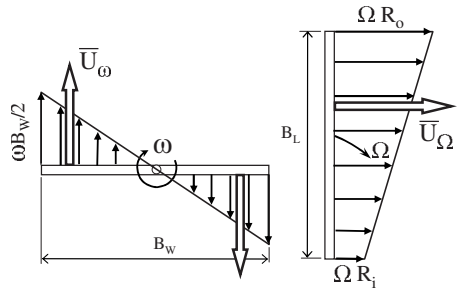


FIG. 8. Blade velocity distributions from the blade pitch modulation frequency ω and the fan rotation Ω.

authors are not aware of cases where the blade pitch is dynamically modulated to produce a desired output signal. Fans have been shown to generate discrete tones as the blades encounter ingested turbulence leading to uneven pressure distributions across the blade,[12–15] which is analogous to the controlled fluctuations of the rotary speaker. Helicopter rotors conspicuously generate discrete sound from turbulence or vortex interactions,[13,16] though the main rotors are not baffled and the blade pitch is not modulated at a signal frequency.

An in-depth analysis of the radiation might employ the Ffowcs Williams–Hawkings extension of Lighthill's acoustic analogy applied to sources in motion,[16,17] or, evaluation of Kirchhoff's integral for moving surfaces.[18,19] In the case that the blades can be considered point sources in the low-frequency limit, the method of Morfey and Tanna could be applied.[20] All of these methods require significant resource investments for successful application. Instead, we will use Rayleigh's integral based on an effective particle velocity to estimate pressure distributions of the projector.

Considering the baffled fan as a planar radiator producing an effective annular surface normal velocity $U_n(S)$, a Green's function form of the Rayleigh integral may be used to represent the pressure at the field point $\mathbf{x}$ produced by the velocity distribution over the surface S.

$$p(\mathbf{x}) = \frac{i\omega\rho}{2\pi}\int U_n(S)\frac{e^{jkr}}{r}dS, \tag{2}$$

where ρ is the atmospheric density and r the magnitude of the position vector from the center of the fan to the field point.

With the fan rotational frequency Ω greater than the blade modulation frequency ω_s, we assume that the net result of forcing on the surrounding air is production of coherent annular rings of pressure pulses. Since the values of ka are small, the source is acoustically compact and an effective particle velocity representation is justified. The effective particle velocity of the ring is modeled by assuming superposition of the blade rotational velocity, U_Ω, with the particle velocity U_ω induced by the dynamically adjusted blade pitch.

Figure 8 depicts velocity distributions of the blade along the transverse and longitudinal directions. The velocity along the length of the blade (dimension B_L in Figs. 1 and 8) is $U_\Omega(R)=\Omega(R_i+R)$, where R is the radial distance from the hub center and takes values in the range $[R_i,R_o]$. An effective velocity evaluated at the centroid of the distribution

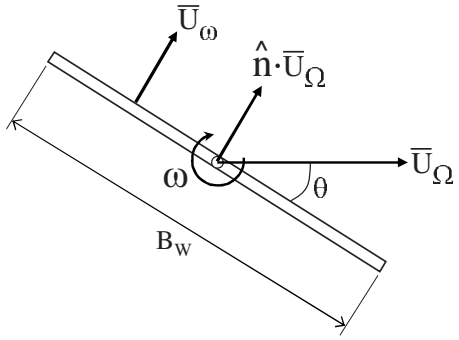


FIG. 9. Effective particle velocity normal to the blade is modeled as the sum of the normal component of fan rotational velocity $\hat{n}\bar{U}_\Omega$ with the effective velocity from the blade pitch modulation $\bar{U}_\omega$.

gives $\bar{U}_\Omega = \Omega[R_i + 2(R_o - R_i)/3]$. Across the width of the blade (dimension B_W) the velocity is ωR_w, with $R_w \in [0, \pm B_W/2]$, and an effective value can be specified as $\bar{U}_\omega = \pm \omega 2B_W/3$.

As the blade varies its angle of attack (θ) according to the signal $s(t)$, the normal component of U_Ω will vary as $\hat{n}\bar{U}_\Omega = \sin(\theta)\bar{U}_\Omega$, as shown in Fig. 9. The composite particle velocity normal to the blade is then

$$\bar{U}_n = \bar{U}_\omega + \hat{n}\bar{U}_\Omega. \tag{3}$$

The maximal value of $\hat{n}\bar{U}_\Omega$ will correspond to the largest excursion of blade pitch $\theta = \theta_{\max}$ during application of the signal $s(t)$. Since we are interested in pressure amplitude we will evaluate $\hat{n}\bar{U}_\Omega$ at $\theta_{\max}$ with the assumption that the maximal particle velocity is the primary contributor to the pressure maxima.

Equation (3) is used in Eq. (2) to compute the radiated pressure of the fan. Equation (2) is numerically evaluated over a grid covering the annular area between the fan hub and baffle. The grid consists of 732 elements of area 2.32 cm^2 each.

The only variable of the model that is not precisely determined is $\theta_{\max}$. This value is directly dependent on the amplitude of the input signal V_{in}. While we were able to measure V_{in}, we did not have access to instrumentation that could quantify θ under operating conditions. Observation of the blades indicated that typical values of $\theta_{\max}$ varied between 25° and 60° with V_{in} in the range of 2–11 V_{rms}. Based on this observation we assume a linear relation between V_{in} and $\theta_{\max}$ with values listed in Table I.

V. RESULTS AND DISCUSSION

A. Model results

Results of Eq. (2) compared to measured data at a fan rotational frequency of $F = 17$ Hz are shown in Table II. Model results are within 2 dB of measured data at drive frequencies of $f_s = 2$, 4, and 8 Hz. However, at 16 Hz the pressure is overestimated by 11 dB. The measured data follow the trend of the transfer function such that the output at $f_s = 16$ Hz is about 3 dB down from the output at $f_s = 8$ Hz, while the model predicts a 6 dB increase. This could indicate that the blade actuation component of the particle velocity $\bar{U}_\omega$ is higher in the model than in the physical system or that the linear combination of the blade rotation and actuation velocities is no longer a valid representation of the particle velocities.

TABLE II. Model results for a fan speed of 17 Hz (1020 rpm) compared to measured values (SPL shown in Fig. 5) at a range of 1 m. Data and model SPL are decibels re Pa^2/Hz.

f_s (Hz)	V_{in}	$\theta_{\max}$ (deg)	Ω/ω_s	$\bar{U}_\Omega/\bar{U}_\omega$	SPL	Model
2	11	58.8	8.5	57.5	18.3	17.7
4	11	58.8	4.3	28.7	25.1	23.8
8	10	55.0	2.1	13.7	28.1	29.8
16	10	55.0	1.1	6.9	25.6	36.4

Examination of the ratio of fan rotation frequency to blade actuation frequency, Ω/ω_s, and of $\bar{U}_\Omega/\bar{U}_{\omega_s}$, reveals that the model breaks down when $\Omega/\omega_s < 2$ or $\bar{U}_\Omega/\bar{U}_\omega < 10$. This indicates that our assumption for model validity that the fan rotational frequency Ω should be greater than the blade modulation frequency ω_s to produce coherent pressure pulses that can be modeled with an effective velocity distribution was off by a factor of 2.

Model results compared to measured data with $F = 11$ Hz are presented in Table III. At signal frequencies of $f_s = 1$ and 4 Hz the predicted and measured values are within 5 dB; however, at the higher frequencies the pressures are severely overpredicted. These results follow the conditions observed above where the model breaks down for $\Omega/\omega_s < 2$ or $\bar{U}_\Omega/\bar{U}_\omega < 10$.

B. Discussion

We have decomposed the effective particle velocity of the rotary speaker into two components, the blade rotational velocity, U_Ω, and the signal modulated blade pitch velocity U_ω. Comparison of the magnitude of these components reveals that the electric motor that drives the hub and blades at frequency Ω is the primary accelerator of the particle velocity. For example, at $F = 17$ Hz and $f_s = 8$ Hz the normal component of effective particle velocity from the hub rotation is $\bar{U}_\Omega = 12.3$ m/s while the blade actuation velocity is $\bar{U}_\omega = 1.3$ m/s. The majority of the consumed power is used to drive the motor and rotate the hub, on the order of 100 W, while the power consumed by the signal actuating the blades is roughly 10 W. This relatively small value of signal power

TABLE I. Blade pitch angles $\theta_{\max}$ in degrees as a function of rms input signal voltage V_{in}

V_{in}	2	3	4	5	6	7	8	9	10	11
$\theta_{\max}$	25.0	28.8	32.5	36.3	40.0	43.8	47.5	51.3	55.0	58.8

TABLE III. Model results for a fan speed of 11 Hz (660 rmp) compared to measured values (SPL shown in Fig. 4) at a range of 1 m. Data and model SPL are decibels re Pa^2/Hz.

f_s (Hz)	V_{in}	θ_{max} (deg)	Ω/ω	$\bar{V}_\Omega/\bar{V}_\omega$	SPL	Model
1	2	25.0	11	37.2	2.9	1.9
1	8	47.5	11	65.4	10.3	6.7
4	4	32.5	2.8	12.0	11.9	16.6
4	6	40.0	2.8	14.1	17.7	18.0
8	5	36.3	1.4	6.6	10.2	23.9
8	7	43.8	1.4	7.6	15.4	25.1
16	6	40.0	0.7	3.6	5.7	31.6
16	8	47.5	0.7	4.1	10.0	32.4

producing high levels of acoustic power results in large transfer function amplitudes and good transduction efficiency.

Measurements of radiated sound power found SPLs at a range of 1 m of 28 dB re Pa^2/Hz (122 dB re 20 $\mu Pa^2/Hz$). It was clearly demonstrated that increasing the value of rotor velocity or blade pitch actuation produced increases in SPL. This is consistent with the expectation that increased particle velocities are a contributing factor to radiation performance. Measured values of SPL at a range of 1820 m found significant signal-to-noise ratios, indicating that detection ranges of the projected infrasound could exceed 2 km. Unexpectedly high values of SPL were recorded at 12 and 16 Hz at a range of 1820 m. It is suggested that the building which forms the back volume of the speaker is resonating at these drive frequencies, which may increase the farfield SPL. Another candidate mechanism for the increased levels are changes in atmospheric propagation conditions. Further study is warranted to research this discrepancy.

A coarse measurement of angular response found reasonable agreement with monopole radiation as expected for a baffled source with $kR_o \ll 1$. However, a roughly 2 dB reduction in the on-axis response was observed at all frequencies. Use of the Rayleigh integral model indicates that radiation from the building wall may be contributing to this, though reduction in SPL of 2 dB could not be reproduced. This represents another area where additional research is needed.

A simple model based on effective particle velocities applied to Rayleigh's integral was developed to predict radiated pressures. The model assumes that the velocity distribution can be considered a linear superposition of effective blade velocities from the fan rotation and the blade pitch actuation. Model results at $F=17$ Hz were within 2 dB of measured values at signal frequencies of $f_s=2,4,8$ Hz, but overpredicted pressure by 11 dB at $f_s=16$ Hz. Model values at $F=11$ Hz were within 5 dB at $f_s=1$ and 4 Hz, but were too large by more than 10 dB for $f_s=8$ and 16 Hz. It was observed that when $\Omega/\omega_s<2$ the model results are poor. This may be an indication that the effective particle velocity representation is no longer valid. A detailed model with improved particle velocity distributions and analysis methods are suggested to address this problem.

VI. CONCLUSION

The rotary subwoofer is a novel acoustic transducer capable of efficiently generating controllable infrasound signals at high SPLs. It is suggested that the speaker overcomes the canonical difficulty of conventional transducers operating in the infrasound band, that the radiation resistance is near zero with radiation impedance dominated by mass-loading, by changing the impedance to fluid displacement dominated by virtue of increased particle velocities. Another unique aspect of the rotary subwoofer is the high transduction efficiency in terms of input signal power to radiated sound power.

The development of a controllable infrasound generator will likely find utility in several acoustic applications. For example, the field calibration of infrasound arrays is operated by the International Monitoring System of the Comprehensive Test Ban Treaty Organization, and as a tool for assessing atmospheric propagation studies. Other applications might include the field simulation of geophysical or anthropogenic infrasound sources to assist in the development of infrasound sensors and detection algorithms.

ACKNOWLEDGMENTS

We wish to acknowledge the support of the National Consortium for Measures and Signatures Intelligence (MASINT) Research (NCMR).

[1]B. J. Bedard and T. M. Georges, "Atmospheric infrasound," Phys. Today **53**(3), 32–37 (2000).
[2]S. P. Neill, "Acoustical characteristics of a very low frequency sound source," MS thesis, University of Mississippi, Oxford, MS (1993).
[3]K. Walker, "A portable infrasonic sensor calibrator down to at least 8 Hz," 30th Monitoring Research Review on Ground-Based Nuclear Explosion Monitoring Technologies, Portsmouth, VA, September 23–25, 2008; http://www.nemre.nnsa.doe.gov/prod/researchreview/2008/PAPERS/06-08.PDF (Last viewed November, 2008).
[4]N. J. Alton, Jr., "Fluid damped acoustic enclosure system," U.S. Patent No. 5,281,777 (25 January 1994).
[5]J. K. Domen, "Infrasonic Helmholtz resonator," U.S. Patent No. 6,665,413 (16 December 2003).
[6]Eminent Technologies, 225 East Palmer Street, Tallahassee, FL 32301, http://www.eminent-tech.com/main.html (Last viewed June, 2008).
[7]J. Park and M. Garcés, "A rotary subwoofer as an infrasonic source," 2007 Infrasound Technology Workshop, Tokyo, Japan, November 13–16, 2007.
[8]Chaparral Physics, Geophysical Institute, University of Alaska, P.O. Box 757320, Fairbanks, AK 99775-7320, http://www.gi.alaska.edu/chaparral/ (Last viewed June, 2008).
[9]L. C. Sutherland and H. E. Bass, "Atmospheric absorption at high altitudes," J. Acoust. Soc. Am. **115**, 1012–1032 (2004).
[10]D. Fee and M. Garcés, "Infrasonic tremor in the diffraction zone," Geophys. Res. Lett. **34**, L16826 (2007).
[11]L. E. Kinsler, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982).
[12]R. Mani, "Noise due to interaction of inlet turbulence with isolated stators and rotors," J. Sound Vib. **17**, 251–260 (1971).
[13]S. E. Wright, "Discrete radiation from rotating periodic sources," J. Sound Vib. **17**, 437–498 (1971).
[14]W.-S. Chiu, G. C. Lauchle, and D. E. Thompson, "Subsonic axial flow fan noise and unsteady rotor force," J. Acoust. Soc. Am. **85**, 641–647 (1989).
[15]S. J. Majumdar and N. Peake, "Noise generation by the interaction between ingested turbulence and a rotating fan," J. Fluid Mech. **359**, 181–216 (1998).
[16]K. S. Brentner and F. Farassat, "Modeling aerodynamically generated

sound of helicopter rotors," Prog. Aerosp. Sci. **39**, 83–102 (2003).
[17]J. E. Ffowcs Williams and D. L. Hawkings, "Sound generation by turbulence and surfaces in arbitrary motion," Philos. Trans. R. Soc. London, Ser. A **264**, 321–342 (1969).
[18]F. Farassat and M. K. Myers, "Extension of Kirchoff's formula to radiation from moving surfaces," J. Sound Vib. **123**, 451–459 (1988).
[19]M. K. Myers and J. S. Hausmann, "On the application of the Kirchoff formula for moving surfaces," J. Sound Vib. **139**, 174–178 (1990).
[20]C. L. Morfey and H. K. Tanna, "Sound radiation from a point force in circular motion," J. Sound Vib. **15**, 325–351 (1971).

A low-noise differential microphone inspired by the ears of the parasitoid fly *Ormia ochracea*

R. N. Miles, Q. Su, W. Cui, and M. Shetye
Department of Mechanical Engineering, State University of New York, P.O. Box 6000, Binghamton, New York 13902-6000

F. L. Degertekin, B. Bicen, and C. Garcia
G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 801 Ferst Drive Northwest, Atlanta, Georgia 30332-0405

S. Jones
Department of Mechanical Engineering, State University of New York, P.O. Box 6000, Binghamton, New York 13902-6000

N. Hall
G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 801 Ferst Drive Northwest, Atlanta, Georgia 30332-0405

(Received 22 May 2008; revised 4 November 2008; accepted 25 January 2009)

A miniature differential microphone is described having a low-noise floor. The sensitivity of a differential microphone suffers as the distance between the two pressure sensing locations decreases, resulting in an increase in the input sound pressure-referred noise floor. In the microphone described here, both the diaphragm thermal noise and the electronic noise are minimized by a combination of novel diaphragm design and the use of low-noise optical sensing that has been integrated into the microphone package. The differential microphone diaphragm measures 1×2 mm^2 and is fabricated out of polycrystalline silicon. The diaphragm design is based on the coupled directionally sensitive ears of the fly *Ormia ochracea*. The sound pressure input-referred noise floor of this miniature differential microphone has been measured to be less than 36 dBA. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3082118]

PACS number(s): 43.38.Kb, 43.38.Ne, 43.66.Ts [AJZ] Pages: 2013–2026

I. INTRODUCTION

Any pressure-sensitive microphone that has an output that depends on the direction of a propagating sound wave must detect the difference in acoustic pressure at a minimum of two points in space. While directional microphones have been used in audio applications for many decades, they can be very challenging to design for applications that have substantial size constraints. One important application for miniature directional microphones is in hearing aids. It is well-known that the use of directional acoustic sensing in hearing aids can be a very effective means of reducing the influence of unwanted background acoustic noise and as a result can significantly improve the understanding of speech in noisy environments (Amlani *et al.*, 2006; Blamey *et al.*, 2006; Hornsby and Ricketts, 2007; Ricketts *et al.*, 2003; Walden *et al.*, 2004). In hearing aids, it is highly desirable that the distance between the points at which the pressure is sensed be kept to a bare minimum, on the order of a few millimeters in order for the size of the device to be cosmetically acceptable. Unfortunately, as this separation distance is reduced, the difference in the sensed pressures is also reduced proportionally. In miniature directional microphones, the desired pressure difference signal can be similar in level to the internal noise and can be very difficult to accurately extract from the average, or common-mode pressure. The main purpose of the present study is to describe a miniature differential microphone that is able to accurately detect pressure differences with minimal influence of microphone noise.

A. Biological inspiration

The approach to the design of the differential microphone described here is based on the mechanical principles employed in the ears of the fly *Ormia ochracea* (Miles *et al.*, 1995). This fly detects spatial gradients in the sound pressure because it has a mechanical connection between the tympana, resulting in the two ears moving in opposite directions in response to pressure differences on their outer surfaces. This differs from the detection of pressure differences by simply allowing the incident sound to drive either side of a single membrane or by subtracting the outputs of two nondirectional microphones. It is well-known that by combining the detection of the pressure gradient along with the pressure, it is straightforward to construct an output that has any of a number of first-order directivity patterns (Olson, 1947; Beranek, 1954). The essence of the mechanical response of *Ormia*'s ears has been shown to be a mechanical realization of a simple sum and difference circuit invented by Blumlein in the earliest days of stereo audio (Blumlein, 1931; Gerzon, 1994; Miles *et al.*, 1997). A simple system to mimic this is described in Miles *et al.* (1997). The mechanical connection between the *Ormia*'s ears has been shown to dramatically

improve the accuracy of estimates of sound source orientation by an analysis of the Cramer–Rao bound (Akcakaya and Nehorai, 2008).

While there are a number of remarkable characteristics of *Ormia*'s auditory system, the feature that the present study focuses on is the fly's unusual apparatus for detecting sound pressure gradients. This is driven by the fact that accurately detecting pressure gradients is the greatest challenge in creating miniature directional microphones. *Ormia*'s ears respond to a combination of the pressure gradient and the pressure to produce directionally sensitive tympanal response (Miles *et al.*, 1995). Achieving the proper combination of the responses due to pressure and pressure gradient is remarkable in a biological system but in a man-made device, it is a simple matter to combine these signals by either an analog circuit or digital processor.

In *Ormia ochracea*'s auditory system, it has been shown that the mechanical structure of the ears allows the response to pressure gradient and pressure to combine in such a way that the tympanum that is closest to a sound source responds with 10–20 dB greater amplitude than the opposite tympanum. This enhanced response amplitude causes the neural sensory cells in the ear closest to the sound source to fire with dramatically less latency than those of the opposite ear (Oshinsky, 1998; Mason *et al.*, 2001). The reduced latency of the neural response provides the central nervous system with the essential information on the orientation of the sound source.

The auditory system of *Ormia ochracea* thus amounts to a multi-stage processor to indicate the direction of a sound source. Because the two tympana are so close together that temporal cues in the incident sound wave are not usable by the nervous system, the mechanical structure of the ears acts to convert these miniscule time of arrival differences into significant differences in tympanal displacements. These differences in displacement are then converted by the sensory nervous system into significantly enhanced temporal differences in neural response, which are then much more readily processed by the central nervous system. The end result for the central nervous system is essentially the expansion or amplification of temporal cues in the incident sound.

Again, the focus of the present effort is to borrow ideas from the fly to help solve the challenging problem in creating miniature directional microphones: the detection of sound pressure gradients.

B. Detection of pressure gradients

The typical method of creating a directional sound sensor is to use a pair of closely spaced nondirectional microphones, as is commonly employed in directional hearing aids. In the absence of microphone noise and mismatches in phase or amplitude sensitivity of the two microphones, it is possible to determine the direction of arrival of a sound relative to the line connecting the microphones with arbitrary accuracy with this first-order small aperture array. Unfortunately, for microphone spacings that are very small relative to the sound wavelength, microphone noise and mismatches in sensitivity or phase can adversely affect the directional response. The primary difficulty is that for small microphone spacings it is difficult to accurately detect differences in the pressures incident on the two microphones. The focus of the present effort is to create a small sensor that provides an accurate measure of the pressure gradient without suffering from these problems.

Any pressure-sensitive diaphragm is a differential sensor because it responds to the difference in pressure on its two sides. Great care is typically taken to convert it to a nondirectional microphone by designing an enclosure to prevent sound from driving one of the two sides of the diaphragm. While a pressure difference sensor is simple to create, the challenge is to obtain adequate response when the size of the device is reduced. Low sensitivity to the desired signal results in a greater influence of electronic and mechanical/thermal noise. In a system that is intended to localize sound, the Cramer–Rao bound on the accuracy of the localization is strongly determined by the system noise floor (see, for example, Friedlander, 1984). In a first-order sound localization system that relies on processing of a combination of the first-order pressure gradient and the pressure [as is done in *Ormia*'s ears (Miles *et al.*, 1997)], the challenge then is to obtain an accurate, low-noise measurement of the pressure gradient; the pressure itself can be measured with much less difficulty because it is much higher in level than the pressure gradient.

The system used for detection of pressure gradients in *Ormia's* ears suggests an alternative method of constructing a differential microphone than the use of a simple single diaphragm. A primary aim of the present study is to examine whether or not this approach can result in a differential microphone having better noise performance than can be achieved with existing approaches. The noise performance of a microphone is strongly dependent on its size. In the case of a differential microphone, the noise performance also strongly depends on the effective distance between the points of pressure measurement. In the following, we have attempted to take these two critical features into account.

C. Other studies of *Ormia*-inspired microphones

A number of researchers have reported the construction of microphones based on the ears of *O. ochracea*. A system is described in Miles *et al.* (1997) that combines the outputs of two nondirectional microphones in a manner that is analogous to the in-phase and out-of-phase vibrational modes of *Ormia's* coupled ears, which provide measures of the pressure and pressure gradient, respectively. This use of the combination of the average and differential pressures to achieve a directionally-sensitive output, as mentioned above, is described in Blumlein (1931). Microphone diaphragms having a directional response based on *Ormia's* ears, using the combination of these two modes of vibration, are described in Yoo *et al.* (2002). The diaphragms were fabricated out of either polycrystalline silicon or Parylene. The diaphragms had dimensions of 1×2 mm^2 and had thicknesses of 1.2 and 2.4 μm for the polysilicon and Parylene diaphragms, respectively. The design and analysis of the acoustic response of these microphones is described in Gibbons and Miles (2000).

Ormia-inspired differential microphone diaphragms are described in Miles *et al.* (2002) having improved sensitivity over those of the earlier design of Gibbons and Miles (2000) and Yoo *et al.* (2002). Measured frequency responses are shown to be in excellent agreement with predictions of the response to incident sound. The microphone diaphragms again had dimensions of 1×2 mm^2 and were fabricated out of polycrystalline silicon. These diaphragms have also been described in Miles *et al.* (2004) and Miles and Hoy (2006) and are predecessors of the design described in the present study. Saito *et al.* (2002) reported a microphone diaphragm inspired by *Ormia's* ears that is supported on a gimbal. This circular diaphragm is made of bronze and has a radius of 10.8 mm and a thickness of 30 μm. A signal processing scheme, intended to be used with a miniature, silicon version of this diaphragm to achieve two-dimensional sound source separation is described in Ono *et al.* (2005).

An *Ormia*-inspired differential microphone having dimensions of 1×2 mm^2, fabricated out of polysilicon and that incorporates optical sensing to obtain an electronic output, was reported by Cui *et al.* (2006). The electronic output of the microphone is shown to have a directivity pattern similar to the figure-eight pattern expected for a first-order differential microphone. A CMOS-compatible directional microphone inspired by *Ormia's* ears is presented in Sung *et al.* (2007). The diaphragm dimensions are 840×410 μm^2. The acoustic response of the diaphragm was detected using a laser vibrometer and was shown to have a directivity pattern similar to the expected figure eight. A pair of mylar diaphragms having radii of 3.5 mm and 22 μm thickness that are glued to a 2.54 cm $\times$ 1.2 mm $\times$ 100 μm rod made of either steel or brass to mimic *Ormia's* ears has been reported by Chen and Yu (2008). It is shown that the incorporation of the connecting rod increased the time delay of the acoustic response of the two diaphragms.

It should also be pointed out that *Ormia's* ability to localize sound has inspired a number of researchers to develop signal processing algorithms for sound source localization that utilize the detection of sound pressure gradients in air (see, for example, Ando, 1995; Stanacevic and Cauwenberghs, 2005; Lockwood and Jones, 2006; Ando *et al.*, 2007; and Mohan *et al.*, 2008).

The differential microphone diaphragm described here is designed to behave like a rigid plate that is supported on carefully-designed flexible pivots (Miles *et al.*, 2004). Pressure gradients acting on the exposed surface of the diaphragm produce a net moment and hence cause it to rotate about its central hinge supports. By taking advantage of microfabrication technology, the diaphragm can be designed to be very lightweight and have very high compliance, enabling it to respond well to minute sound pressure gradients even though the largest dimension of the differential diaphragm is only 2 mm.

D. Optical sensing

In order to transduce the motion of the diaphragm into an electronic signal, we have incorporated a high-sensitivity, low-noise optical sensing scheme providing minimum detectable displacement on the order of 10 fm/$\sqrt{\text{Hz}}$, using approximately 61 μW of laser power (Hall and Degertekin, 2002; Lee *et al.*, 2004; Cui *et al.*, 2006; Miles and Degertekin, 2007). The majority of successful miniature microphone designs employ capacitive sensing, which suffers from instability when used on highly compliant diaphragms because the required bias voltage causes a significant force that attracts the diaphragm toward the biasing electrode. The use of low-noise optical sensing allows our differential microphone diaphragm to have very high compliance, which is important in achieving good sensitivity to pressure gradients.

The use of optical sensing and a highly compliant and responsive diaphragm has resulted in the achievement of very low electronic noise from the microphone. As a result, the noise floor of the microphone is primarily influenced by its thermal noise, resulting from random impacts between the diaphragm and the surrounding air. It is well-known that the equivalent input sound pressure associated with this thermal noise is proportional to the amount of passive energy dissipation in the system (Gabrielson, 1993; Thompson *et al.*, 2002; Hall *et al.*, 2007). The design developed here does not require the use of a backplate electrode, which is typically the dominant cause of viscous damping in capacitive microphones (Thompson *et al.*, 2002; Homentcovschi and Miles, 2004; Homentcovschi and Miles, 2005; Hall *et al.*, 2007). In addition, because the dominant motion of the diaphragm consists of rotation about its flexible supports, acoustic radiation damping and other sources of viscous damping can be kept to a minimum. As a result, the overall design approach leads to a system having low noise relative to the desired pressure gradient being sensed.

In the following, the designs of the microphone diaphragm and the optical sensing scheme are described along with an analysis of the noise performance. The microphone's response and noise outputs are shown to agree very well with predictions. Finally, the noise floor of the differential microphone is compared with what can be achieved with current, commercially available hearing aid microphones.

II. DIFFERENTIAL MICROPHONE DIAPHRAGM AND OPTICAL SENSOR

Figure 1 shows a design model of the microphone diaphragm. It consists of a stiffened plate supported by flexible pivots. The overall dimensions of the diaphragm are 1×2 mm^2. The diaphragm is fabricated out of polycrystalline silicon having a thickness of approximately 1 μm. Stiffeners, indicated in Fig. 1, are incorporated to cause the diaphragm to respond as a rigid plate that is supported on flexible pivots. The stiffeners are approximately 20 μm tall and 2 μm thick. The finite element design model has been used in a numerical optimization procedure to optimize the dimensions of the diaphragm and the stiffeners to achieve low noise and high acoustic sensitivity. Approximate expressions are used for the noise performance and the sensitivity (as derived below), which are used to construct an objective function in the iterative design. The ANSYS finite element model accounts for all of the resonant vibrational modes that can have an influence on the acoustic response. The struc-

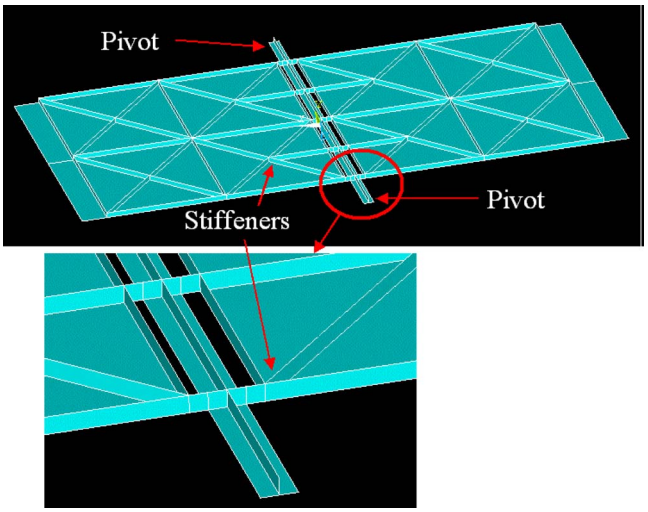


FIG. 1. (Color online) Design model of the microphone diaphragm. The diaphragm consists of a stiffened plate that is supported by flexible pivots. The stiffeners are also indicated in Fig. 4.

tural optimization has been carried out to ensure that the response is dominated by the mode in which the diaphragm rotates about its supports as a rigid plate. The dimensions of the stiffeners have been optimized to attenuate the influence of these undesirable modes. The mode shapes predicted by the design model are shown in Fig. 2. The first mode, consisting of out-of-phase motion of the two ends of the diaphragm, has a predicted resonant frequency of 821.7 Hz while the second mode, having both ends move in-phase, has a predicted resonant frequency of 15 002.7 Hz. The stiffening effect of the air volume within the package behind the microphone diaphragm has been taken into account in the

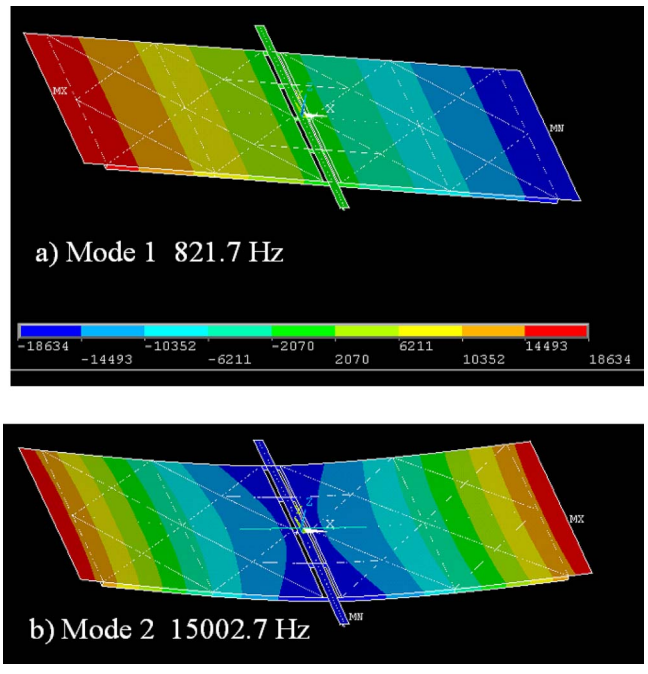


FIG. 2. (Color online) The mode shapes of the first two resonant modes of the diaphragm as predicted by the finite element (ANSYS) design model. The first mode shown in panel (a) consists of the diaphragm rotating as a rigid plate about its supporting pivots. The second mode shown in panel (b) consists of both ends of the diaphragm moving in-phase.

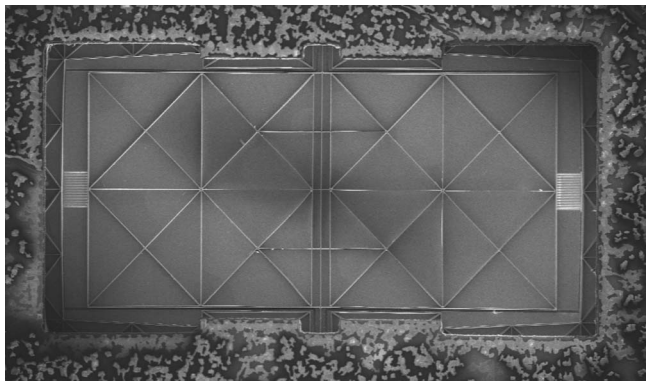


FIG. 3. Scanning electron micrograph of the microphone diaphragm. This image shows the back side of the diaphragm through the hole etched through the back of the wafer.

finite element model. Figure 3 shows a scanning electron micrograph showing the stiffeners as seen through a hole that has been etched through the thickness of the silicon wafer, as described below. Figures 4(a) and 4(b) show a photograph and the mask design of the top surface of the microphone diaphragm. The figure shows gold mirrors at each end of the diaphragm used in the optical sensing system described below.

Figure 5 shows the fabrication process flow for the microphone diaphragm. The fabrication starts with a deep reactive ion trench etch into the 4 in. test grade silicon wafer that acts as the mold for the polysilicon stiffeners. This is followed by a wet oxidation at 1100 °C to grow a 1 μm thick thermal oxide layer on the wafer surface and in the trenches [Fig. 5(a)]. This oxide is used as an etch stop for a subsequent backside cavity etch. The next step is to deposit and planarize polycrystalline silicon to form a 1 μm thick diaphragm having stiffeners. The phosphorus-doped polysilicon is deposited at 580 °C and subsequently annealed at 1100 °C in argon gas for 60 min in order to reduce intrinsic stress in the film. The polysilicon is then etched to define the optical grating and the slits that separate the diaphragm from

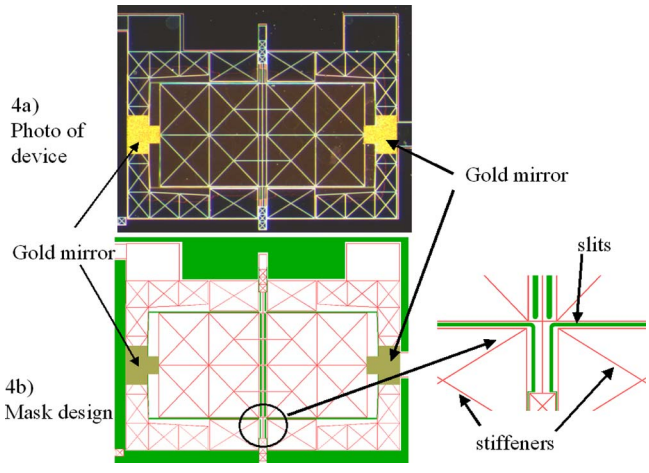


FIG. 4. (Color online) Photograph (a) and the mask design (b) of the top surface of the microphone diaphragm. The figure shows gold mirrors at each end of the diaphragm used in the optical sensing system. A close-up of the mask design is shown to indicate the stiffeners and the slit that is cut to form the perimeter of the diaphragm.

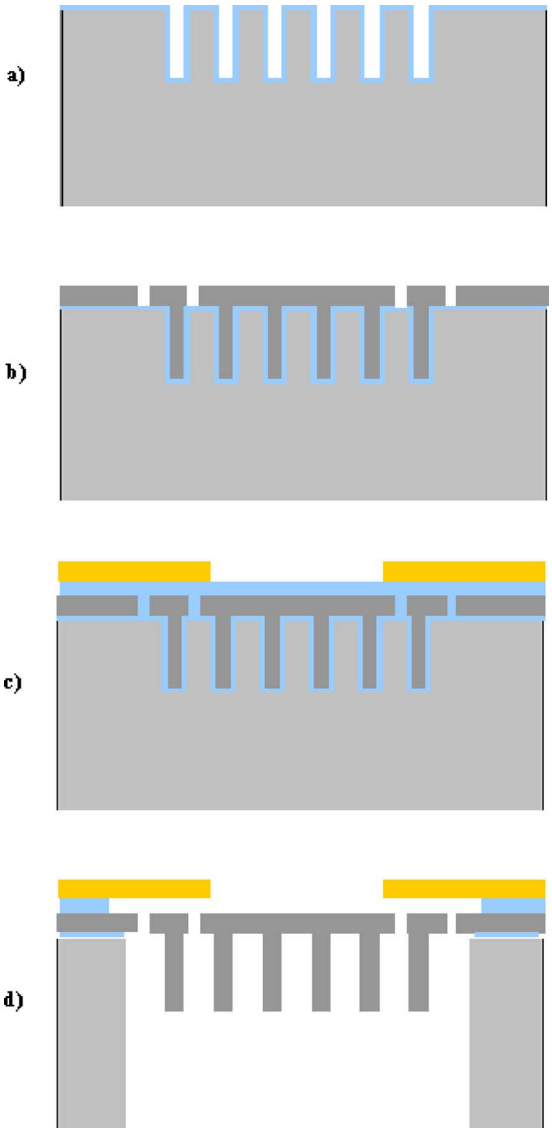


FIG. 5. (Color online) Fabrication process flow for the microphone diaphragm. The fabrication starts with a deep reactive ion trench etch into the 4-in. test grade silicon wafer. This is followed by a wet oxidation to grow a 1 μm thick thermal oxide layer on the wafer surface and in the trenches [Fig. 5(a)]. The next step is to deposit and planarize polycrystalline silicon to form a 1 μm thick diaphragm having stiffeners. The polysilicon is then etched to define the optical grating and the slits that separate the diaphragm from the substrate [Fig. 5(b) and in the inset of Fig. 4(b)]. This is followed by PECVD oxide deposition, chrome and gold seed layer evaporation, and gold plating to form the gold mirrors [Fig. 5(c)]. The back cavity is then etched using a deep reactive ion etch and the thermal and PECVD oxide layers are removed in buffered oxide etchant to release the structures [Fig. 5(d)].

the substrate [Fig. 5(b)]. This is followed by plasma enhanced chemical vapor deposition (PECVD) oxide deposition, chrome and gold seed layer evaporation, and gold plating to form the gold mirrors. Both the oxide layer and the gold mirrors are 5 μm in thickness [Fig. 5(c)]. The back cavity is then etched using a deep reactive ion etch and the thermal and PECVD oxide layers are removed in buffered oxide etchant to release the structures [Fig. 5(d)].

A schematic of the optical sensing scheme is shown in Fig. 6 (Cui *et al.*, 2006; Miles and Degertekin, 2007). This consists of an optical grating (an array of slits), a mirror

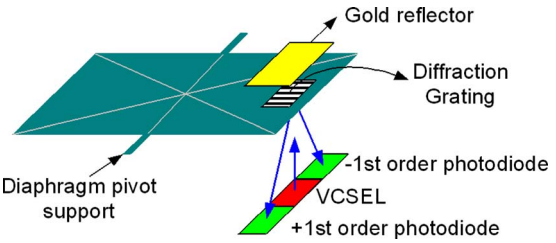


FIG. 6. (Color online) Schematic of the differential microphone diaphragm and the optical components. This consists of an optical grating (an array of slits), a mirror placed above the diaphragm, a light source (in this case a VCSEL), and photodetectors. The light that is reflected back to the photodetectors is modulated by the relative motion between the grating on the diaphragm and the mirror. The electrical output of the photodetectors is thus proportional to the diaphragm motion, providing the sensitivity of an interferometer.

placed above the diaphragm, a light source [in this case a vertical cavity surface emitting laser (VCSEL)], and photodetectors. The light that is reflected back to the photodetectors is modulated by the relative motion between the grating on the diaphragm and the mirror. The electrical output of the photodetectors is thus proportional to the diaphragm motion, providing the sensitivity of an interferometer. Similar optical detection structures have already been used to implement low-noise micromachined omnidirectional microphones (Hall *et al.*, 2005; 2007). Figure 7 shows an end-view of the diaphragm and optical components. The optical components and the silicon chip containing the microphone diaphragm have been assembled in a package depicted in Figs. 8 and 9. The silicon chip containing the directional microphone diaphragm shown in Figs. 8 and 9 contains an array of three microphone diaphragms, two differential diaphragms, and one nondirectional microphone diaphragm. This microphone array on the chip has been fabricated in order to explore the use of co-located directional microphones in a signal processing scheme as part of this effort described in Mohan *et al.* (2008).

In Sec. III, we examine the physical principles that determine the noise performance of the microphone. Experimental results will then be described and compared with predictions.

III. ANALYTICAL MODEL FOR MICROPHONE RESPONSE AND NOISE FLOOR

In order to construct a simplified model that can predict the microphone's noise performance, it is necessary to write down expressions for the response to sound. Because the

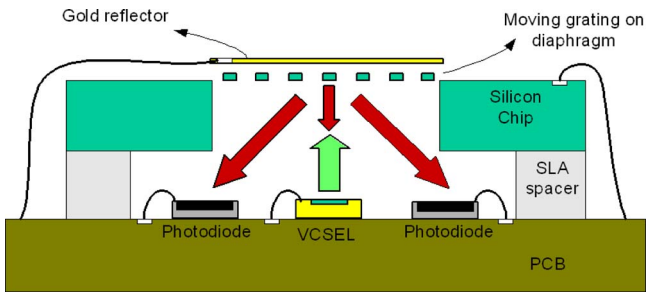


FIG. 7. (Color online) End-view of the diaphragm and optical components.

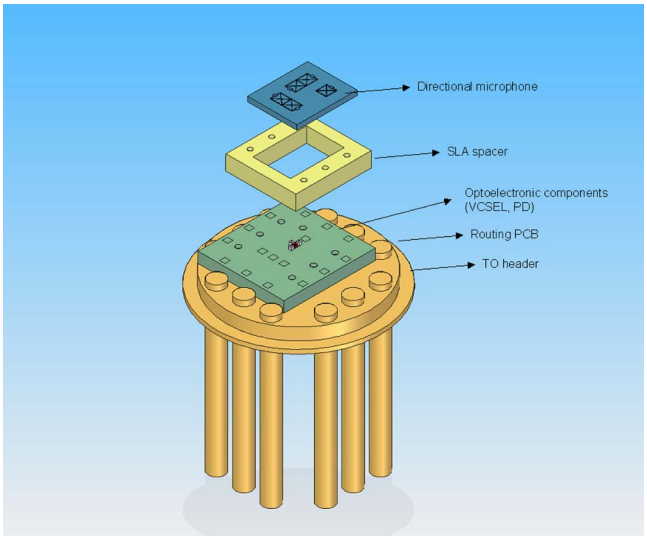


FIG. 8. (Color online) Package assembly design for the optical components and the silicon chip containing the microphone diaphragm.

diaphragm is intended to rotate about its hinge axis due to the sound pressure gradient, we first write the relation between the pressure gradient and the applied moment about the axis of rotation. A simple lumped-parameter model for the diaphragm's rotation can then be written in terms of the equivalent mass moment of inertia, I, about the axis of rotation, the equivalent torsional stiffness, k_t, and the equivalent viscous damping constant, c_t. The equivalent driving moment due to thermal excitation is then determined along with the equivalent pressure of an incident plane sound wave that would cause the same response as the thermal excitation. While the microphone diaphragm has been designed to respond primarily in its first resonant mode, the second mode of vibration, in which the two ends translate in-phase, can influence the response. The effects of the response of this translational mode along with the noise due to electronic circuits are included in an expression for the equivalent sound pressure noise.

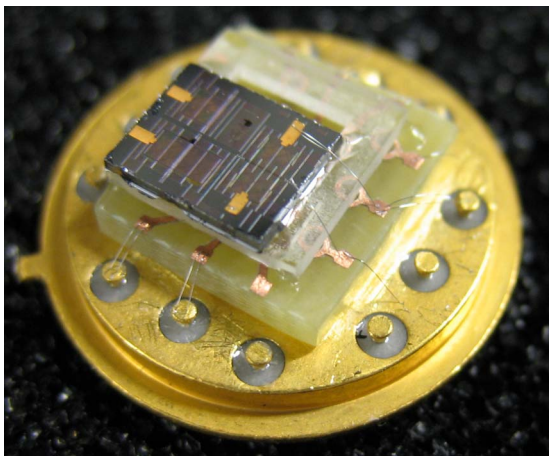


FIG. 9. (Color online) Photograph of the assembled microphone. The optical components are below the microphone diaphragm chip and are not visible.

A. Relation between sound pressure gradient and driving moment

We will assume that the microphone diaphragm is small relative to the wavelength of sound so that the spatial variation in pressure is small over its surface. Taking the origin of our coordinate system to be at the center of the diaphragm with the rotation oriented in the y direction and the length L measured in the x direction, the sound pressure as a function of x and t can be expanded in a two-term Taylor's series,

$$p(x,t) \approx p(0,t) + x \left.\frac{\partial p}{\partial x}\right|_{x=0}. \tag{1}$$

The moment applied by the sound field about the diaphragm's axis of rotation is then given by

$$M(t) = \int_{-L/2}^{L/2} bxp(x,t)dx \approx \int_{-L/2}^{L/2} bxp(0,t)dx + \int_{-L/2}^{L/2} bx^2 \left.\frac{\partial p}{\partial x}\right|_{x=0} dx, \tag{2}$$

where we have used the expansion in Eq. (1). b is the width of the diaphragm. Carrying out the integrations on the right in Eq. (2) gives

$$M(t) \approx \left.\frac{\partial p}{\partial x}\right|_{x=0} I_A, \tag{3}$$

where, for this simple rectangular diaphragm, the area moment of inertia about the axis of rotation is $I_A = bL^3/12$.

Equation (3) shows that the moment applied to the diaphragm is the product of the pressure gradient in the x direction and the area moment of inertia.

As mentioned above, the diaphragm has been designed so that the only vibrational modes that influence its motion in the audible frequency range have one of the two mode shapes shown in Fig. 2. While additional, ultrasonic modes could easily be included in our noise analysis, our present aim is to construct a model containing only the essential features that guide design.

B. Lumped-parameter model for diaphragm rotation

We first consider the motion due to the mode which is dominated by rotation about the pivot axis. Then, the contribution to the noise due to the second mode of vibration will be accounted for. Since the diaphragm has been designed so that its dominant vibrational mode consists of pure rotation as a rigid body, the equation of motion of the diaphragm in terms of its rotation, θ, may be taken to be

$$I\ddot{\theta} + k_t\theta + c_t\dot{\theta} = M(t), \tag{4}$$

where I is the mass moment of inertia, k_t is the equivalent torsional stiffness, and c_t is the effective dashpot constant for rotation. Equation (4) can also be expressed in the more convenient form

$$\ddot{\theta} + \omega_0^2\theta + 2\omega_0\zeta\dot{\theta} = \frac{M(t)}{I}, \tag{5}$$

where $\omega_0^2 = k_t/I$, and ζ is the damping ratio, $\zeta = c_t/2\sqrt{k_t I}$.

Because our aim is to examine the response of the diaphragm due to random thermal excitation from the surrounding medium, and compare that to the response due to sound, the incident sound field will be taken to be a weakly stationary random signal. This enables us to examine the effects of both the acoustic and thermal excitations using power spectral densities. By using Eq. (5), the power spectral density of the rotation of the diaphragm can be shown to be

$$S_{\theta\theta}(\omega) = \frac{S_{MM}(\omega)/I^2}{(\omega_0^2 - \omega^2)^2 + (2\omega_0\zeta\omega)^2}, \tag{6}$$

where $S_{MM}(\omega)$ is the power spectral density of the applied moment, $M(t)$.

C. Response due to thermal excitation

We would like to compare the response of the microphone due to acoustic excitation with that due to random thermal excitation from the surrounding gas. The equipartition theorem and the Nyquist relation can be used to calculate the response due to thermal excitation. Assuming, for now, that the response is dominated by a single rotational mode of vibration as assumed above, then

$$\tfrac{1}{2}K_B T = \tfrac{1}{2}k_t E[\theta^2]_n, \tag{7}$$

where $K_B = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and T is the absolute temperature. $E[\theta^2]_n$ is the mean square response due to the thermal excitation. If the power spectral density of the moment due to thermal excitation, S_{MM}^N, is independent of frequency, ω, then for a system governed by Eq. (5), the mean square response of the rotation is obtained by integrating the power spectral density of Eq. (6) over all frequencies (Lin, 1967),

$$E[\theta^2]_n = \frac{\pi S_{MM}^N}{2\zeta\omega_0^3 I^2} = \frac{\pi S_{MM}^N}{k_t c_t} = \frac{K_B T}{k_t}, \tag{8}$$

where we have also used Eq. (7). Equation (8) then gives the two-sided power spectral density of the applied moment due to the thermal excitation (with units of $(\mathrm{N\,m})^2/\mathrm{rad/s}$),

$$S_{MM}^N = \frac{K_B T c_t}{\pi}. \tag{9}$$

D. Equivalent input pressure due to a plane sound wave

Since directional microphones are typically characterized in terms of their response to incident plane waves, it is convenient to re-cast Eq. (9) to determine the equivalent sound pressure of an incident wave that would correspond to the response due to thermal excitation. If the sound pressure is due to a plane wave incident with a speed c_0 at an angle ϕ relative to the direction that is tangent to the plane of the diaphragm, then the pressure on the diaphragm can be expressed as

$$p(x,t) = p\left(t - \frac{x}{c_0}\cos(\phi)\right). \tag{10}$$

In the case where the sound pressure is a weakly stationary random signal, the power spectral density of the sound pressure at $x=0$ may be obtained from the Fourier transform of the autocorrelation function of the pressure, $R_{pp}(\tau) = E[p(t)p(t-\tau)]$, where E denotes the expected value,

$$S_{pp}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\omega\tau} R_{pp}(\tau)\, d\tau. \tag{11}$$

Since the moment, $M(t)$, in Eq. (3) depends on the pressure gradient, we need the power spectral density of the gradient to determine the power spectral density of the response. From Eq. (10), the gradient of the pressure, $dp/dx = p_x(x,t)$ at $x=0$ is

$$p_x(t) = \dot{p}(t)\frac{-\cos(\phi)}{c_0}. \tag{12}$$

The autocorrelation function of the gradient, $\partial p/\partial x$ at $x=0$, may then be written as

$$R_{p_x p_x}(\tau) = E[p_x(t)p_x(t-\tau)] = \left(\frac{\cos(\phi)}{c_0}\right)^2 E[\dot{p}(t)\dot{p}(t-\tau)]. \tag{13}$$

Taking the transform of Eq. (13) as in Eq. (11) enables us to obtain the power spectral density of the pressure gradient as

$$S_{p_x p_x}(\omega) = \left(\frac{\omega\cos(\phi)}{c_0}\right)^2 S_{pp}(\omega). \tag{14}$$

By using Eqs. (3) and (14), we can obtain the relation between the power spectral density of the applied moment and that of the pressure for this plane wave,

$$S_{MM}(\omega) = I_A^2\left(\frac{\omega\cos(\phi)}{c_0}\right)^2 S_{pp}(\omega). \tag{15}$$

Equations (6) and (15) give the power spectrum of the response as

$$S_{\theta\theta}(\omega) = \frac{(\omega\cos(\phi)/c_0)^2 (I_A/I)^2 S_{pp}(\omega)}{(\omega_0^2 - \omega^2)^2 + (2\omega_0\omega\zeta)^2}. \tag{16}$$

The relation between the spectral densities in Eq. (15) along with Eq. (9) can allow us to write the power spectral density of the pressure due to a plane wave, $S_{pp}(\omega)$, that would cause the same response as the thermal excitation,

$$S_{pp}^N(\omega) = \frac{S_{MM}^N}{I_A^2(\omega\cos(\phi)/c_0)^2} = \frac{K_B T c_t}{\pi I_A^2(\omega\cos(\phi)/c_0)^2}. \tag{17}$$

This result is normally evaluated for sound that is incident at $\phi=0$, the most sensitive direction. The power spectrum of the response due to thermal excitation may be written by replacing $S_{pp}(\omega)$ with $S_{pp}^N(\omega)$ in Eq. (16),

$$S^N_{\theta\theta}(\omega) = \frac{(\omega\cos(\phi)/c_0)^2(I_A/I)^2 S^N_{pp}(\omega)}{(\omega_0^2-\omega^2)^2+(2\omega_0\omega\zeta)^2} = \frac{K_B T c_t/(\pi I^2)}{(\omega_0^2-\omega^2)^2+(2\omega_0\omega\zeta)^2}. \quad (18)$$

E. Influence of the translational mode on noise performance

While the motion of the diaphragm is typically dominated by rotation, θ, at most frequencies of interest in audio applications, it is also important to account for the second mode shown in Fig. 2(b) in which the ends of the diaphragm move in the direction normal to the plane. This translational mode will be described by the coordinate $X(t)$. Assuming that the properties of the diaphragm are symmetric about its axis of rotation, the coordinates $X(t)$ and $\theta(t)$ will be uncoupled. The equation governing $X(t)$ is

$$m\ddot{X} + kX + c\dot{X} = p(0,t)A,$$

$$\ddot{X} + \omega_X^2 X + 2\omega_X\zeta_X\dot{X} = p(0,t)A/m, \quad (19)$$

where m is the effective mass, k is the equivalent stiffness, c is the equivalent dashpot constant, ω_X is the natural frequency, and ζ_X is the damping ratio. A is the equivalent area of the diaphragm and $p(0,t)$ is the surface average of the sound pressure acting on the diaphragm [which is also equal to $p(x,t)$ at $x=0$].

As in Eq. (6), the power spectral density of the translational mode is related to that of the acoustic pressure by

$$S_{XX}(\omega) = \frac{S_{pp}(\omega)(A/m)^2}{(\omega_X^2-\omega^2)^2+(2\omega_X\zeta_X\omega)^2}. \quad (20)$$

If both the rotational and translational modes are accounted for, the equipartition theorem states that each uncoupled mode receives the same thermal energy, $K_BT/2$, so that the thermal energy imparted to the translational mode is

$$\tfrac{1}{2}K_BT = \tfrac{1}{2}kE[X^2]_N, \quad (21)$$

where $E[X^2]_N$ is the mean square of the response of the mode due to thermal excitation. Proceeding as in Eqs. (5), (6), and (18), the power spectral density of the translational mode due to thermal excitation is

$$S^N_{XX}(\omega) = \frac{K_BTc/(\pi m^2)}{(\omega_X^2-\omega^2)^2+(2\omega_X\omega\zeta_X)^2}. \quad (22)$$

Because the response is detected at the end of the diaphragm at a distance d from the pivot, the detected displacement, $W(t)$, is related to the translation and rotation by

$$W(t) = X(t) + d\theta(t). \quad (23)$$

F. Equivalent input pressure noise including effects of electronic circuit noise

Since the output of the microphone must be detected using non-ideal electronic circuits, there will be some contribution to the noise due to noise generated in the circuits. This electronic noise can be represented by an equivalent displacement noise of the diaphragm $W^e(t)$, having a power spectral density given by $S^{Ne}_{WW}(\omega)$. The noise due to the electronics is not examined in detail here; the relative contribution due to laser intensity noise, shot noise in the photodetectors, and noise from other electronic components will be examined in other studies.

Since the thermal excitation is comprised of spatially random impacts of gas molecules with the diaphragm, the electronic noise and the thermal responses of the two modes are uncorrelated. Equations (18), (22), and (23) allow us to write the power spectral density of the thermal response of $W(t)$ as

$$S^N_{WW}(\omega) = \frac{K_BTc_td^2/(\pi I^2)}{(\omega_0^2-\omega^2)^2+(2\omega_0\omega\zeta)^2} + \frac{K_BTc/(\pi m^2)}{(\omega_X^2-\omega^2)^2+(2\omega_X\omega\zeta_X)^2} + S^{Ne}_{WW}(\omega). \quad (24)$$

Having the power spectral density of the equivalent displacement noise in Eq. (24), it is possible to estimate the power spectral density of the sound pressure that would produce this level of noise. To compute this input pressure-referred spectral density, note that the power spectral density of the displacement response due to sound pressure is

$$S_{WW}(\omega) = |H_{PW}(\omega)|^2 S_{pp}(\omega), \quad (25)$$

where

$$H_{PW}(\omega) = \frac{A/m}{\omega_X^2-\omega^2+2\omega_X\omega\zeta_X\hat{J}} + \frac{\hat{J}\omega\cos(\phi)dI_A/(c_0I)}{\omega_0^2-\omega^2+2\omega_0\omega\zeta\hat{J}}. \quad (26)$$

Equations (24) and (25) may be used to obtain an equivalent sound pressure power spectral density that would cause the predicted noise response power spectrum, $S^N_{WW}(\omega)$,

$$S^{Nt}_{pp}(\omega) = \left(\frac{K_BTc_td^2/(\pi I^2)}{(\omega_0^2-\omega^2)^2+(2\omega_0\omega\zeta)^2} + \frac{K_BTc/(\pi m^2)}{(\omega_X^2-\omega^2)^2+(2\omega_X\omega\zeta_X)^2} + S^{Ne}_{WW}(\omega)\right)\frac{1}{|H_{PW}(\omega)|^2}. \quad (27)$$

In order to allow direct comparisons with measured results, the predicted spectral density of the displacement noise in Eq. (24) can be modified to predict the output voltage noise, $S^N_{VV}(\omega)$, by

$$S^N_{VV}(\omega) = |H_{WV}(\omega)|^2 \times S^N_{WW}(\omega), \quad (28)$$

where $H_{WV}(\omega)$ is the transfer function between the diaphragm displacement and the output voltage in V/m, given in Table I.

Knowing the displacement relative to the pressure, $H_{PW}(\omega)$, in Eq. (26) and the output voltage for a given displacement, $H_{WV}(\omega)$, the output voltage relative to the input sound pressure is

TABLE I. Parameters used in predictions.

Estimated diaphragm mechanical parameters
Mass moment of inertia, $I=5\times10^{-15}$ kg m^2
Rotation mode natural frequency, $f_r=735$ Hz
Rotational mode damping ratio, $\xi_r=0.16$
Rotational dashpot constant$=c_t=7.389\times10^{-12}$ N m s/rad
Translational mode effective mass, $m=4\times10^{-8}$ kg
Translational mode natural frequency, $f_t=15220$ Hz
Translational mode damping ratio, $\xi_t=0.25$
Translational mode dashpot constant$=c=0.001\,912\,6$ N s/m
Diaphragm dimensions
Length, $1=2.2\times10^{-3}$ m
Width, $b=10^{-3}$ m
Area moment of inertia, $I_A=bl^3/12=8.87\times10^{-13}$ m^4
Electrical parameters
Electrical/mechanical sensitivity: $H_{WV}(f)=0.0056$ V/nm
Parameters for noise prediction
Boltzmann's constant$=K_B=1.38\times10^{-23}$ J/K
Absolute temperature$=T=293$ K
Diaphragm parameters from ANSYS finite element design model
$I=3.49\times10^{-15}$ kg m^2
Rotational mode natural frequency, $f_r=821.7$ Hz
Translational mode natural frequency, $f_t=15\,002.7$ Hz

$$H_{PV}(\omega)=H_{PW}(\omega)\times H_{WV}(\omega). \tag{29}$$

IV. COMPARISONS WITH EXISTING TECHNOLOGY

A straightforward and common way to create a directional acoustic sensor is through the use of a pair of omnidirectional microphones. The equivalent input-referred noise of this directional system can be estimated knowing the noise floors of the two-omnidirectional microphones, $S_{oo}^N(\omega)$, and by accounting for the loss of sensitivity to sound when the outputs of the microphones are subtracted to achieve a directional response. This loss of sensitivity can be accounted for by multiplying the output sensitivity of the omnidirectional microphones by $\omega d_o/c$, where ω is the frequency in rad/s d_o is the distance between the microphones, and c is the sound speed. In addition, because the power of the uncorrelated noise of the microphones is added when the output signals are subtracted, the resultant noise power spectrum is effectively increased by a factor of 2. By taking the difference of the two signals to achieve a directional output, the equivalent sound input-referred noise is then

$$S_{dd}^N(\omega)=2\left(\frac{c}{\omega d_o}\right)^2 S_{oo}^N(\omega). \tag{30}$$

To compare the equivalent input pressure noise levels of a two-omnidirectional microphone system with that of the present study, it can be more convenient to express the equivalent input noise levels in decibels. If $\mathrm{SPL}_{oo}(\omega)$ is the equivalent input sound pressure level of the two-omnidirectional microphones, then the input-referred noise level of the directional system may be estimated from

$$\mathrm{SPL}_{dd}(\omega)=10\log_{10}S_{dd}^N(\omega)=\mathrm{SPL}_{oo}(\omega)+3-20\log_{10}\left(\frac{\omega d_o}{c}\right). \tag{31}$$

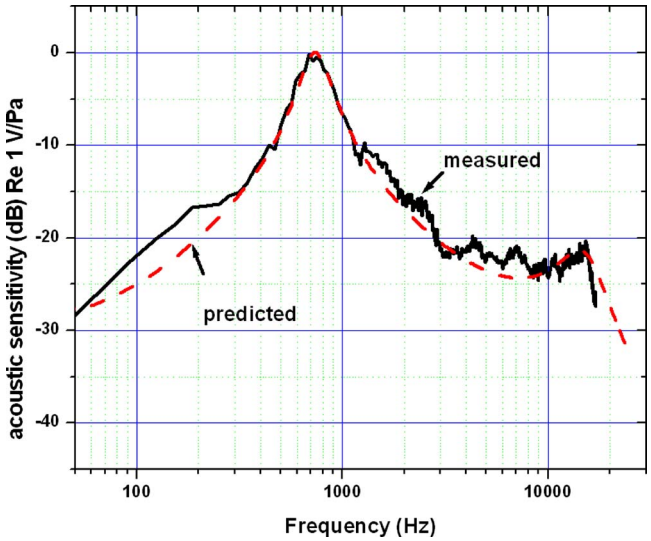


FIG. 10. (Color online) Comparison of predicted and measured electrical output in response to sound. The results are normalized relative to the sound pressure that is incident on the diaphragm. The data shown in this and the following figures were obtained by averaging auto- and/or cross power spectral density estimates from 20 records of the measured signals. The spectra were computed using the fast Fourier transform algorithm applied to the discrete time domain data. A coherence squared function was also calculated between the measured input signal and the response. Data were rejected when the coherence was less than 0.85. The coherence was typically greater than 0.95.

V. MEASURED RESULTS

The electrical output of the microphone shown in Fig. 9 has been measured relative to the incident sound pressure as measured by a calibrated reference microphone. The results are shown as a function of frequency in hertz in Fig. 10 along with those predicted by Eq. (29). The sound source was oriented in the most sensitive direction for the microphone. The measured and predicted results are in very close agreement. Figure 10 shows that the frequency dependence of the sensitivity is that of a bandpass filter as expected from the form of Eq. (16) for the power spectral density of the diaphragm rotation.

Since a flat frequency response is sought in most microphone applications, it may be desirable to incorporate a compensation filter that will result in a flat frequency response over most of the audible range. This frequency compensation can be accomplished without influencing the sound input-referred noise because, as will be discussed below, the noise is dominated by the diaphragm thermal noise (and not the noise from the electronics). The diaphragm thermal noise is superimposed on the detected sound pressure so its relation to the sound pressure is not affected by subsequent process-

ing of the signal. In addition, an active electronic feedback system is being developed that provides the beneficial effects of damping for both the transient response and frequency response of the diaphragm without adversely affecting the noise performance (Miles, 2008).

The physical parameters used in evaluating Eq. (29) to obtain the predicted results shown in Fig. 10 are given in Table I. The mass moment of inertia, I, the natural frequencies, equivalent mass, m, and the damping constants have been identified empirically through a least squares curve fit of the measured displacement of the diaphragm due to sound using the method outlined in the Appendix. Table I also lists the mass moment of inertia and the first two natural frequencies predicted by our finite element design model. The predicted and measured frequencies of the rotational mode are 821.7 and 735 Hz, respectively (12% error). The predicted and measured in-phase mode frequencies are 15002.7 and 15220 Hz, respectively (1.4% error). The resonant frequencies of the fabricated device were thus very close to those of our finite element design model.

The agreement between measured and predicted results indicates that the measured characteristics of the diaphragm are consistent with those expected from an analytical model based only on the physical dimensions and known material properties. It is also important to note that the frequencies of the first two modes are separated by roughly a factor of 20, which results in the rotational mode dominating the response over a wide range of frequencies.

The relation between the output voltage and the diaphragm displacement, $H_{WV}(\omega)$ in Eqs. (28) and (29), has been obtained by measuring the electrical output and the diaphragm displacement (measured using a Veeco optical profilometer) due to a voltage applied between the mirror shown in Figs. 4, 6, and 7 and the diaphragm.

The measured directivity pattern of the diaphragm is shown in Fig. 11 along with the figure-eight pattern expected for an ideal differential microphone. The measured results are shown to be in reasonable agreement with the ideal dipole pattern.

The output noise of the microphone as predicted by Eq. (28), in terms of the single-sided power spectral density (with units of $\mathrm{V}/\sqrt{\mathrm{Hz}}$), is compared to the measured output noise in Fig. 12. The measurements were performed while the microphone was placed in an anechoic chamber at the Georgia Institute of Technology to minimize sources of acoustic noise. The noise floor of the anechoic chamber was measured using a 1/2 in. Larson Davis model 2541 microphone and is significantly below the measured noise shown here; external acoustic noise does not influence these results except at very low frequencies. For frequencies below about 3 kHz, the output noise is strongly influenced by the thermal noise of the rotational mode of the diaphragm as predicted by the first term in Eq. (24). As shown in Fig. 12, the total microphone noise is dominated by the thermal noise for frequencies below about 3 kHz. The discrepancy between the predicted and measured noise below 100 Hz is likely due to acoustic noise in the anechoic chamber. The discrepancy at frequencies above 3 kHz is probably due to electronic noise in the optical system.

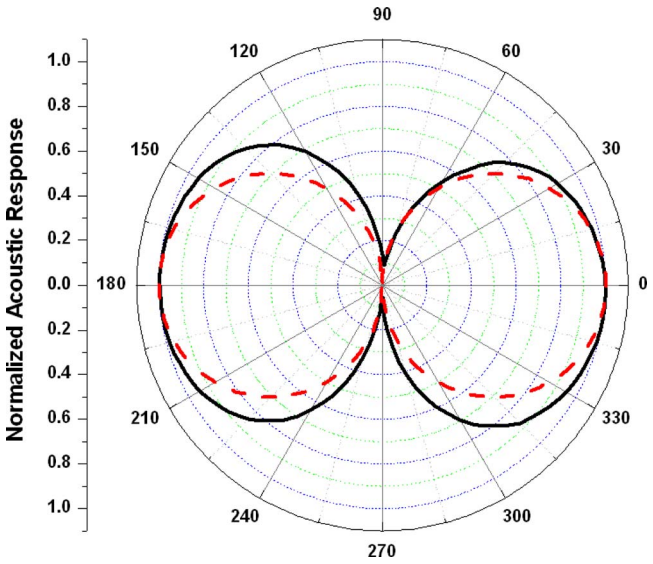


FIG. 11. (Color online) The measured directivity pattern of the diaphragm along with the figure eight pattern expected for an ideal differential microphone. The directivity measurements are performed in an anechoic chamber. The directional microphone is mounted on a rotation stage which can be controlled with a motion controller. The sound source is at a fixed location and driven by a pure tone sine wave at a frequency of approximately 800 Hz.

The equivalent sound input pressure-referred noise as predicted by Eq. (27) is compared to the measured input pressure-referred noise in Fig. 13. The measured results are obtained by dividing those of Fig. 12 by those given in Fig. 10.

The one-third octave band levels of the narrowband input pressure-referred noise of Fig. 13 are shown in Fig. 14. The corresponding dBA levels of the measured and predicted data are 35.6 and 34 dBA, respectively. The difference between the predicted and measured results is most likely due to electronic noise $S_{WW}^{Ne}(\omega)$, which has been neglected in the predictions and has not been measured independently.

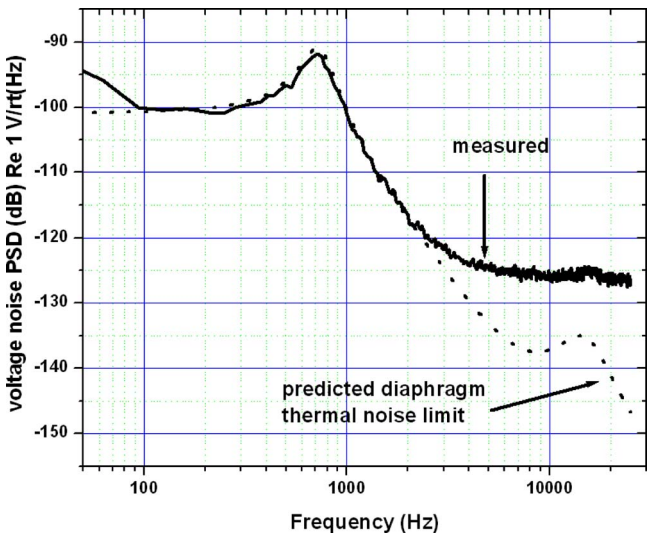


FIG. 12. (Color online) The output noise of the microphone as predicted by Eq. (28) compared to the measured output noise. The predicted results were obtained while neglecting the contribution due to the electronic noise, $S_{WW}^{Ne}(\omega)$, in Eq. (24).

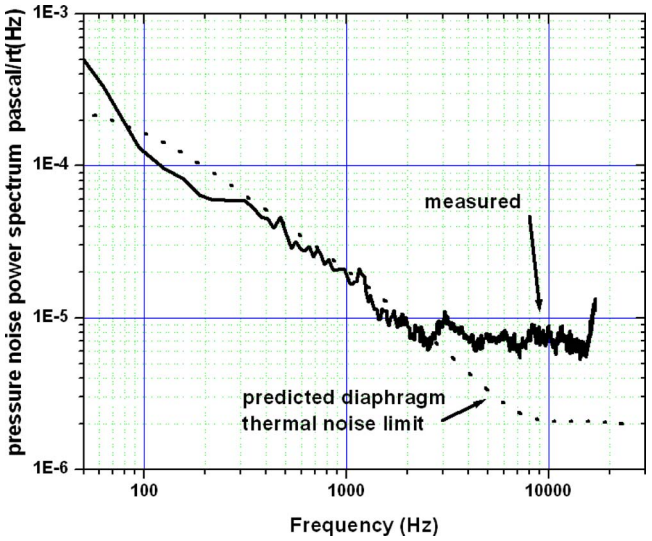


FIG. 13. (Color online) The equivalent sound input pressure-referred noise as predicted by Eq. (27) compared to the measured input pressure-referred noise. The measured results are obtained by dividing those of Fig. 12 by those given in Fig. 10. As in Fig. 12, the thermal noise limit was estimated by neglecting the contribution of electronic noise.

To compare the results obtained in the present study with those of commercially available hearing aid microphones, Eq. (31) has been evaluated to estimate the equivalent input noise of a differential sound sensor composed of two Knowles low-noise EM omnidirectional microphones. The noise floor of these omnidirectional microphones, $\mathrm{SPL}_{oo}(\omega)$, obtained from Knowles.com, is shown in the figure along with the results of Eq. (31). It is assumed that the spacing between the microphones is $d_o=10$ mm. The estimated noise of the differential sound sensor constructed by subtracting the output of the Knowles EM microphones is substantially higher than that of the differential microphone described here. In the microphone developed here, one can roughly estimate the effective separation distance between the pressure sensors to be the distance between the centers of each side, separated by the axis of symmetry, which is half the diaphragm length, or 1 mm. It is important to note that the effective separation distance, d_o, of the omnidirectional microphones is ten times that of the microphone developed in the present study. As can be seen in Eq. (31), this factor of 10 gives the two microphone system a noise floor advantage of approximately 20 dB, yet its noise is substantially higher over much of the audible frequency range and its A-weighted noise floor of 47.9 dBA is more than 10 dB higher.

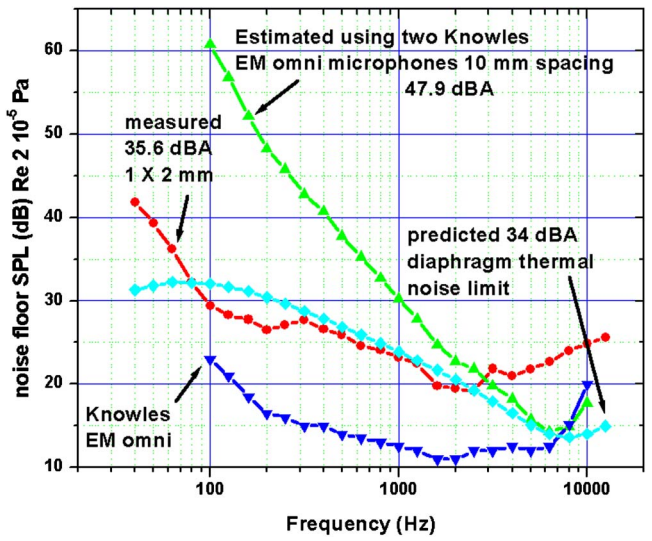


FIG. 14. (Color online) The one-third octave band levels of the narrowband input pressure-referred noise of Fig. 13. Equation (31) has been evaluated to estimate the equivalent input noise of a differential sound sensor composed of two Knowles low-noise EM omnidirectional microphones. The noise floor of these omnidirectional microphones, $\mathrm{SPL}_{oo}(\omega)$, obtained from Knowles.com is shown in the figure along with the results of Eq. (31). It is assumed that the spacing between the microphones is $d_o=10$ mm. The estimated noise of the differential sound sensor constructed by subtracting the output of the Knowles EM microphones is substantially higher than that of the differential microphone described here.

VI. CONCLUSIONS

The miniature differential microphone described in this paper is able to detect pressure gradients with minimal influence of microphone noise. This is accomplished through the use of a microphone diaphragm that rotates about a highly compliant central hinge due to an applied moment resulting from an incident sound wave. By minimizing the mass, stiffness, and passive damping in the diaphragm design, it is possible to achieve adequate sensitivity to sound and low thermal/mechanical noise. The use of a low-noise optical detection scheme has resulted in a design in which the total microphone noise is dominated by the diaphragm thermal/mechanical noise rather than noise in the electronic components over much of the audible frequency range.

ACKNOWLEDGMENTS

This work has been supported by the National Institute on Deafness and Other Communication Disorders Grant No. R01DC005762-05 for a NIH Bioengineering Research Partnership and by Grant No. R01DC009429 to R.N.M. The authors would like to thank Dr. Lynn Luethke of the National Institutes of Health for her support. Support has also been provided by NSF Grant No. ECS-0643055. The authors would like to thank the following researchers for their contributions to this work through many helpful discussions: Ron Hoy, Cornell University; Steve Thompson, Penn State University; Ruth Bentler, University of Iowa; Doug Jones and Michael Lockwood, University of Illinois Urbana Champaign; and Pete Loeppert and Ray Kirchhoefer, Knowles Inc.

APPENDIX: EMPIRICAL ESTIMATION OF DIFFERENTIAL MICROPHONE MECHANICAL PARAMETERS

The predictions of the response and noise performance through the analysis described in Sec. III above rely on knowledge of the effective mass, stiffness, and damping of the microphone diaphragm. For the rotational mode, these are I, k_t, and c_t in Eq. (4). These have been obtained by processing the measured response power spectra for the diaphragms due to a weakly stationary random acoustic excitation. The parameters are identified through a least squares approach where the mean square error in the model equation

[Eq. (4)] is minimized (Su, 2005). The approach leads to a linear system of three normal equations for the three unknown parameters, I, k_t, and c_t. Power spectral density functions estimated from the measured random response are used to obtain the statistical information for construction of the normal equations. The empirical estimation process is outlined in this Appendix. It is shown that this procedure is essentially similar to the frequency domain method that is derived for deterministic signals presented in Raol *et al.* (2004), Napolitano *et al.* (2001), and Morelli (1999).

As discussed in Sec. III above, the mechanical system is modeled as a linear second order resonator, as described in Eq. (4). If the actual measured rotation of the diaphragm, $\theta(t)$, is entered into the equation, errors in the model will introduce an error in the net moment, $\epsilon(t)$, so that Eq. (4) becomes

$$\tilde{I}\ddot{\theta}(t) + \tilde{c}_t\dot{\theta}(t) = \tilde{k}_t\theta(t) = M(t) + \epsilon(t). \tag{A1}$$

The mean squared error, $E[\epsilon^2(t)]$, where $E[\ ]$ is the expected value, is minimized by differentiating with respect to each unknown parameter,

$$\frac{\partial E[\epsilon^2(t)]}{\partial \tilde{I}} = 0,$$

$$\frac{\partial E[\epsilon^2(t)]}{\partial \tilde{c}_t} = 0,$$

$$\frac{\partial E[\epsilon^2(t)]}{\partial \tilde{k}_t} = 0. \tag{A2}$$

Equations (A1) and (A2) lead to a system of equations for the unknown parameters that may be written in matrix form as

$$\begin{bmatrix} E[\ddot{\theta}^2] & E[\dot{\theta}\ddot{\theta}] & E[\theta\ddot{\theta}] \\ E[\ddot{\theta}\dot{\theta}] & E[\dot{\theta}^2] & E[\theta\dot{\theta}] \\ E[\ddot{\theta}\theta] & E[\dot{\theta}\theta] & E[\theta^2] \end{bmatrix} \begin{pmatrix} \tilde{I} \\ \tilde{c}_t \\ \tilde{k}_t \end{pmatrix} = \begin{pmatrix} E[M\ddot{\theta}] \\ E[M\dot{\theta}] \\ E[M\theta] \end{pmatrix}. \tag{A3}$$

The expected values in Eq. (A3) may be expressed in terms of correlation functions $R_{\dot{\theta}\dot{\theta}}(\tau)$ and $R_{M\dot{\theta}}(\tau)$ when $\tau=0$. For weakly stationary signals, the correlation functions may be determined from the inverse Fourier transforms of the auto- and cross spectral density functions, $S_{\dot{\theta}\dot{\theta}}(\omega)$ and $S_{M\dot{\theta}}(\omega)$, by the *Weiner–Khinchine relations*.

$$E[\dot{\theta}(t)\dot{\theta}(t-\tau)] = R_{\dot{\theta}\dot{\theta}}(\tau) = \int_{-\infty}^{\infty} e^{\hat{J}\omega\tau} S_{\dot{\theta}\dot{\theta}}(\omega)d\omega,$$

$$E[M(t)\dot{\theta}(t-\tau)] = R_{M\dot{\theta}}(\tau) = \int_{-\infty}^{\infty} e^{\hat{J}\omega\tau} S_{M\dot{\theta}}(\omega)d\omega. \tag{A4}$$

The two-sided autospectral density function is real valued and an even function of frequency, and the two-sided cross spectral density function is complex with the real part an even function of frequency and the imaginary part an odd function of frequency (Bendat and Piersol, 1986; Lin, 1967). Therefore $S_{\dot{\theta}\dot{\theta}}(\omega)=S_{\dot{\theta}\dot{\theta}}(-\omega)$ and $S_{M\dot{\theta}}(-\omega)=S^*_{M\dot{\theta}}(\omega)$, where * is the complex conjugate, and the expected values in Eq. (A3) are

$$E[\ddot{\theta}^2] = \int_{-\infty}^{\infty} \omega^2 S_{\dot{\theta}\dot{\theta}}(\omega)d\omega = 2\int_0^{\infty} \omega^2 S_{\dot{\theta}\dot{\theta}}(\omega)d\omega,$$

$$E[\dot{\theta}^2] = \int_{-\infty}^{\infty} S_{\dot{\theta}\dot{\theta}}(\omega)d\omega = 2\int_0^{\infty} S_{\dot{\theta}\dot{\theta}}(\omega)d\omega,$$

$$E[\theta^2] = \int_{-\infty}^{\infty} \frac{1}{\omega^2} S_{\dot{\theta}\dot{\theta}}(\omega)d\omega = 2\int_0^{\infty} \frac{1}{\omega^2} S_{\dot{\theta}\dot{\theta}}(\omega)d\omega,$$

$$E[\ddot{\theta}\theta] = E[\theta\ddot{\theta}] = -\int_{-\infty}^{\infty} S_{\dot{\theta}\dot{\theta}}(\omega)d\omega = -2\int_0^{\infty} S_{\dot{\theta}\dot{\theta}}(\omega)d\omega,$$

$$E[\ddot{\theta}\dot{\theta}] = E[\dot{\theta}\ddot{\theta}] = \int_{-\infty}^{\infty} \hat{J}\omega S_{\dot{\theta}\dot{\theta}}(\omega)d\omega = 0,$$

$$E[\dot{\theta}\theta] = E[\theta\dot{\theta}] = -\int_{-\infty}^{\infty} \frac{\hat{J}}{\omega} S_{\dot{\theta}\dot{\theta}}(\omega)d\omega = 0, \tag{A5}$$

and

$$E[M\ddot{\theta}] = \int_{-\infty}^{\infty} \hat{J}\omega S_{M\dot{\theta}}(\omega)d\omega = 2\int_0^{\infty} \Re[\hat{J}\omega S_{M\dot{\theta}}(\omega)]d\omega,$$

$$E[M\dot{\theta}] = \int_{-\infty}^{\infty} S_{M\dot{\theta}}(\omega)d\omega = 2\int_0^{\infty} \Re[S_{M\dot{\theta}}(\omega)]d\omega,$$

$$E[M\theta] = -\int_{-\infty}^{\infty} \frac{\hat{J}}{\omega} S_{M\dot{\theta}}(\omega)d\omega = -2\int_0^{\infty} \Re\left[\frac{\hat{J}}{\omega} S_{M\dot{\theta}}(\omega)\right]d\omega, \tag{A6}$$

where $\Re$ denotes the real part. Solving Eq. (A3) for the mechanical parameters gives

$$\tilde{I} = \frac{E[\theta^2]E[M\ddot{\theta}] - E[\theta\ddot{\theta}]E[M\theta]}{E[\ddot{\theta}^2]E[\theta^2] - E[\theta\ddot{\theta}]^2},$$

$$\tilde{c}_t = \frac{E[M\dot{\theta}]}{E[\dot{\theta}^2]},$$

$$\tilde{k}_t = \frac{E[\ddot{\theta}^2]E[M\theta] - E[\ddot{\theta}\theta]E[M\ddot{\theta}]}{E[\ddot{\theta}^2]E[\theta^2] - E[\theta\ddot{\theta}]^2}. \tag{A7}$$

The finite Fourier transforms of the signals are

$$\dot{\Theta}(\omega,T) = \frac{1}{2\pi}\int_{-T}^{T} e^{-\hat{J}\omega t}\dot{\theta}(t)dt,$$

$$\mathcal{M}(\omega,T) = \frac{1}{2\pi}\int_{-T}^{T} e^{-\hat{J}\omega t} M(t)dt, \tag{A8}$$

which enables the calculation of the auto- and cross spectral densities,

$$S_{\dot{\theta}\dot{\theta}}(\omega) = \lim_{T\to\infty} \frac{\pi}{T} E[|\dot{\Theta}(\omega,T)|^2],$$

$$S_{M\dot{\theta}}(\omega) = \lim_{T\to\infty} \frac{\pi}{T} E[\mathcal{M}^*(\omega,T)\dot{\Theta}(\omega,T)]. \tag{A9}$$

When the signals are deterministic and their Fourier transforms exist, the identification procedure is the same as that described by Raol *et al.* (2004), Napolitano *et al.* (2001), and Morelli (1999). The Fourier transform can be applied to the governing equation and arranged as $\underline{Z}=[Y]\underline{\beta}+\underline{\varepsilon}$ for discrete data:

$$\begin{pmatrix} \mathcal{M}(\omega_1) \\ \mathcal{M}(\omega_2) \\ \vdots \\ \mathcal{M}(\omega_n) \end{pmatrix} = \begin{bmatrix} \ddot{\Theta}(\omega_1) & \dot{\Theta}(\omega_1) & \Theta(\omega_1) \\ \ddot{\Theta}(\omega_2) & \dot{\Theta}(\omega_2) & \Theta(\omega_2) \\ \vdots & \vdots & \vdots \\ \ddot{\Theta}(\omega_n) & \dot{\Theta}(\omega_n) & \Theta(\omega_n) \end{bmatrix} \begin{pmatrix} \tilde{I} \\ \tilde{c}_t \\ \tilde{k}_t \end{pmatrix} + \underline{\varepsilon}. \tag{A10}$$

The least squares solution for complex data is

$$\underline{\beta} = (\Re([Y]^{*T}[Y]))^{-1}\Re([Y]^{*T}\underline{Z}) \tag{A11}$$

and is the same solution as described in terms of the spectral densities when the signals are deterministic with $\ddot{\Theta}(\omega) = \hat{j}\omega\dot{\Theta}(\omega)$ and $\Theta(\omega) = -(\hat{j}/\omega)\dot{\Theta}(\omega)$.

1. Parameter estimation for the differential microphone diaphragm

The power spectra required in the above procedure were estimated from measured acoustic response of the microphones. Using a random excitation signal to drive a loudspeaker, a Doppler laser vibrometer (Polytec PSV-300H) was used to measure the out-of-plane velocity of the diaphragm and a probe microphone (Bruel and Kjaer Type 4182) measured the sound pressure at the diaphragm. Transfer functions of the velocity with respect to the sound pressure were estimated with a spectrum analyzer from averaged estimates of the auto- and cross power spectral densities. Two transfer functions for the diaphragm response were used for the mechanical parameter identification, $H_{p\dot{w}_1}(\omega)$ and $H_{p\dot{w}_2}(\omega)$, where $\dot{w}_1(t)$ and $\dot{w}_2(t)$ are the velocities at two points on opposite halves of the differential diaphragm separated by the pivot, and $p(t)$ is the sound pressure at the pivot. The spectral density functions in Eq. (A6) were estimated from $H_{p\dot{w}_1}(\omega)$, $H_{p\dot{w}_2}(\omega)$, and the sound pressure autospectral density $S_{pp}(\omega)$.

Assuming that the diaphragm is symmetric about the pivot axis, the rotation is approximated as $\theta=(w_1-w_2)/(2d)$, and the driving moment is estimated from the sound pressure using the plane wave model, $M=-\dot{p}(t)\cos(\phi)I_a/c_0$. The auto- and cross spectral density functions that are needed in the identification procedure are related to the measured transfer functions by $S_{\dot{\theta}\dot{\theta}}(\omega) = |H_{p\dot{\theta}}(\omega)|^2 S_{pp}(\omega)$ and $S_{M\dot{\theta}}(\omega) = H^*_{pM}(\omega)H_{p\dot{\theta}}(\omega)S_{pp}(\omega)$, where $H_{p\dot{\theta}}(\omega) = (H_{p\dot{w}_1}(\omega) - H_{p\dot{w}_2}(\omega))/(2d)$, and $H^*_{pM}(\omega) = \hat{j}\omega\cos\phi I_A/c_0$. This gives

$$S_{\dot{\theta}\dot{\theta}}(\omega) = \left|\frac{H_{p\dot{w}_1}(\omega) - H_{p\dot{w}_2}(\omega)}{2d}\right|^2 S_{pp}(\omega),$$

$$S_{M\dot{\theta}}(\omega) = \frac{\hat{J}\omega\cos\phi I_A}{c_0}\left(\frac{H_{p\dot{w}_1}(\omega) - H_{p\dot{w}_2}(\omega)}{2d}\right)S_{pp}(\omega). \tag{A12}$$

The same procedure is used to estimate the parameters of the translational mode, governed by Eq. (19). The average diaphragm displacement is $X=(w_1+w_2)/2$ and the average normal force due to the sound pressure is $F=pA$. The spectral density functions for the identification algorithm in this case are

$$S_{\dot{X}\dot{X}}(\omega) = \left|\frac{H_{p\dot{w}_1}(\omega) + H_{p\dot{w}_2}(\omega)}{2}\right|^2 S_{pp}(\omega),$$

$$S_{F\dot{X}}(\omega) = \frac{A(H_{p\dot{w}_1}(\omega) + H_{p\dot{w}_2}(\omega))}{2} S_{pp}(\omega). \tag{A13}$$

The integrals of the spectral density functions for the expected values are approximated by summation over a user selected frequency range, where the mode in question is dominant, and where data quality is good, as verified by coherence squared functions.

Akcakaya, M., and Nehorai, A. (**2008**). "Performance analysis of the *Ormia ochracea's* coupled ears," J. Acoust. Soc. Am. **124**, 2100–2105.

Amlani, A. M., Rakerd, B., and Punch, J. L. (**2006**). "Speech-clarity judgments of hearing-aid-processed speech in noise: Differing polar patterns and acoustic environments," Int. J. Audiol. **45**, 319–330.

Ando, S. (**1995**). "An intelligent three-dimensional vision sensor with ears," Sens. Mater. **7**, 213–231.

Ando, S., Ono, N., and Fujita, Y. (**2007**). "Partial differential equation-based algorithm of sound source localization with finest granularity in both time and frequency," in Proceedings of International Conference Networked Sensing Systems (INSS2007), pp. 229–234.

Bendat, J. S., and Piersol, A. G. (**1986**). *Random Data Analysis and Measurement Procedures* (Wiley, New York).

Beranek, L. L. (**1954**). *Acoustics* (Acoustical Society of America, Woodbury, NY).

Blamey, P. J., Fiket, H. J., and Steele, B. R. (**2006**). "Improving speech intelligibility in background noise with an adaptive directional microphone," J. Am. Acad. Audiol **17**, 519–530.

Blumlein, A. D. (**1931**). "Directional effect in sound systems," British Patent No. 394,325.

Chen, Z., and Yu, M. (**2008**). "Biology-inspired acoustic sensors for sound source localization," in *Proceedings of the SPIE Smart Structures and Materials & Nondestructive Evaluation and Health Monitoring, 15th Annual International Symposium*, San Diego, CA, Vol. **6932–108**.

Cui, W., Jones, S. A., Miles, R. N., Degertekin, F. L., Hall, N., and Bicen, B. (**2006**). "Optical sensing in a directional MEMS microphone inspired by the ears of the parasitoid fly, *Ormia ochracea*," in *19th IEEE Conference on Micro Electro Mechanical Systems (MEMS 2006)*, Istanbul, Turkey.

Freidlander, B. (**1984**). "On the Cramer–Rao bound for time delay and Doppler estimation," IEEE Trans. Inf. Theory **IT-30**, 575–580.

Gabrielson, T. (**1993**). "Mechanical-thermal noise in micromachined acoustic and vibration sensors," IEEE Trans. Electron Devices **40**, 903–909.

Gerzon, M. A. (**1994**). "Applications of Blumlein shuffling to stereo microphone techniques," J. Audio Eng. Soc. **42**, 435–453.

Gibbons, C., and Miles, R. N. (**2000**). "Design of a biomimetic directional microphone diaphragm," in *Proceedings of the International Mechanical Engineering Congress and Exposition*, Orlando, FL.

Hall, N. A., and Degertekin, F. L. (**2002**). "An integrated optical interferometric detection method for micromachined capacitive acoustic transducers," Appl. Phys. Lett. **80**, 3859–3861.

Hall, N. A., Bicen, B., Lee, W., Jeelani, K., Qureshi, S., Okandan, M., and

Degertekin, F. L. (**2005**). "Micromachined microphones with diffraction-based optical displacement detection," J. Acoust. Soc. Am. **118**, 3000–3009.
Hall, N. A., Okandan, M., Littrell, R., Bicen, B., and Degertekin, F. L. (**2007**). "Micromachined optical microphone structures with low thermal-mechanical noise levels," J. Acoust. Soc. Am. **122**, 2031–2037.
Homentcovschi, D., and Miles, R. N. (**2004**). "Modeling of viscous damping of perforated planar microstructures. Applications in acoustics," J. Acoust. Soc. Am. **116**, 2939–2947.
Homentcovschi, D., and Miles, R. N. (**2005**). "Viscous damping of perforated planar micromechanical structures," Sens. Actuators, A **119**, 544–552.
Hornsby, B. W., and Ricketts, T. A. (**2007**). "Effects of noise source configuration on directional benefit using symmetric and asymmetric directional hearing aid fittings," Ear Hear. **28**, 177–186.
Lee, W., Hall, N. A., Zhou, Z., and Degertekin, F. L. (**2004**). "Fabrication and characterization of a micromachined acoustic sensor with integrated optical readout," IEEE J. Sel. Top. Quantum Electron. **10**, 643–651.
Lin, Y. K. (**1967**). *Probabilistic Theory of Structural Dynamics* (Robert E. Krieger, Malabar, FL).
Lockwood, M. E., and Jones, D. L. (**2006**). "Beamformer performance with acoustic vector sensors in air," J. Acoust. Soc. Am. **119**, 608–619.
Mason, A., Oshinsky, M., and Hoy, R. R. (**2001**). "Hyperacute directional hearing in a microscale auditory system," Nature (London) **410**, 686–690.
Miles, R. N. (**2008**). "Low-noise directional hearing aid microphone using optical sensing with electronic feedback," National Institutes of Health Grant No. 1R01DC009429-01.
Miles, R. N., and Degertekin, F. L. (**2007**). "Optical sensing in a directional MEMS microphone," U.S. Patent No. US2007/0165896 A1.
Miles, R. N., and Hoy, R. R. (**2006**). "The development of a biologically-inspired directional microphone for hearing aids," Audiol. Neuro-Otol. **11**, 86–94.
Miles, R. N., Cui, W., Miller, R. A., Su, Q., Tan, L., and Weinstein, M. G. (**2002**). "Response of a biologically inspired MEMS differential microphone diaphragm," in *Proceedings of the SPIE AeroSense 2000*, Orlando, FL, Paper No. 4743-15.
Miles, R. N., Robert, D., and Hoy, R. R. (**1995**). "Mechanically coupled ears for directional hearing in the parasitoid fly *Ormia ochracea*," J. Acoust. Soc. Am. **98**, 3059–3070.
Miles, R. N., Sundermurthy, S., Gibbons, C., Hoy, R., and Robert, D. (**2004**). "Differential Microphone," U.S. Patent No. 6,788,796 B1.
Miles, R. N., Tieu, T. D., Robert, D., and Hoy, R. R. (**1997**). "A mechanical analysis of the novel ear of the parasitoid fly *Ormia ochracea*," in *Proceedings: Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lion, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 18–24.
Mohan, S., Lockwood, M. E., Kramer, M. L., and Jones, D. L. (**2008**). "Localization of multiple acoustic sources with small arrays using a coherence test," J. Acoust. Soc. Am. **123**, 2136–2147.
Morelli, E. A. (**1999**). "Real-time parameter estimation in the frequency domain," AIAA-99-4043.
Napolitano, M. R., Song, Y., and Seanor, B. (**2001**). "On-line parameter estimation for restructurable flight control systems," Aircraft Design. **4**, 19–50.
Olson, H. F. (**1947**). *Elements of Acoustical Engineering* (D. Van Nostrand Co., Inc., New York), Chap. 8.3.
Ono, N., Arita, T., Senjo, Y., and Ando, S. (**2005**). "Directivity Steering Principle for Biomimicry Silicon Microphone," in *13th International Conference on Solid-State Sensors, Actuators and Microsystems Transducers '05*, Seoul, Korea.
Oshinsky, M. L. (**1998**). "Physiological correlates of sound localization in a parasitoid fly, *Ormia ochracea*," Ph.D. thesis, Cornell University, Ithaca, NY.
Raol, J. R., Girija, G., and Singh, J. (**2004**). *Modeling and Parameter Estimation of Dynamic Systems*, The Institution of Electrical Engineers, London.
Ricketts, T., Henry, P., and Gnewikow, D. (**2003**). "Full time directional versus user selectable microphone modes in hearing aids," Ear Hear. **24**, 424–439.
Saito, A., Ono, N., and Ando, S. (**2002**). "Micro gimbal diaphragm for sound source localization with mimicking Ormia Ochracea," in *Proceedings of the 41st SICE Annual Conference Society for Instrument and Control Engineers (SICE)*.
Stanacevic, M., and Cauwenberghs, G. (**2005**). "Micropower gradient flow acoustic localizer," IEEE Trans. Circuits Syst., I: Regul. Pap. **52**, 2148–2157.
Su, Q. T. (**2005**). "Measurement and characterization of miniature silicon microphone diaphragms," Ph.D. thesis, State University of New York Binghamton, Binghamton, NY.
Sung, P.-H., Chen, J.-Y., Yen, K.-H., and Wu, C.-Y. (**2007**). "CMOS compatible directional microphone," in Microsystems, Packaging, Assembly and Circuits Technology, IMPACT 2007, Vols. **1–3**, pp. 149–152.
Thompson, S. C., LoPresti, J. L., Ring, E. M., Nepomuceno, H. G., Beard, J. J., Ballad, W. J., and Carlson, E. V. (**2002**). "Noise in miniature microphones," J. Acoust. Soc. Am. **111**, 861–886.
Walden, B. E., Surr, R. K., Cord, M. T., and Dyrlund, O. (**2004**). "Predicting hearing aid microphone preference in everyday listening," J. Am. Acad. Audiol **15**, 365–396.
Yoo, K., Yeh, J.-L. A., Tien, N. C., Gibbons, C., Su, Q., and Miles, R. N. (**2002**). "Fabrication of biomimetic 3-D structured diaphragms," Sens. Actuators, A **97–98**, 448–456.

Determination of the random anisotropic elasticity layer using transient wave propagation in a fluid-solid multilayer: Model and experiments

Christophe Desceliers[a] and Christian Soize
Modélisation et Simulation Multi Echelle, Université Paris-Est, MSME FRE3160 CNRS, 5 bd Descartes, 77454 Marne-la-Vallee, France

Quentin Grimal and Maryline Talmant
Laboratoire d'Imagerie Parametrique, Université Paris VI, UMR7623 CNRS, 15 rue de l'école de médecine, 75006 Paris, France

Salah Naili
Laboratoire de Mecanique Physique, Université Paris-Est, B2OA UMR7052 CNRS, 61 avenue du Général de Gaulle, 94010 Creteil Cedex, France

(Received 22 February 2008; revised 6 January 2009; accepted 3 February 2009)

The aim of this paper is to introduce a simplified model for an uncertain solid layer sandwiched between two acoustic fluid layers and using the ultrasonic characterization with an acoustic source placed in one fluid layer. Uncertainties are taken into account with a probabilistic model of the elasticity tensor. Its parameters are the mean value of the random tensor and a dispersion parameter that controls the statistical fluctuation level. The characterization of the solid layer given a database of actual measurements consists in the determination of the (i) elastic parameters of the mean elasticity model, (ii) the dispersion parameter, and (iii) mass density of the solid. This is performed with a numerical solver of wave propagation and for *in vivo* data collected previously. The model is representative of measurements of human bone properties with the so-called axial transmission technique. The capability of the model to predict the velocity of the first experimental arriving signal in the statistical sense is proved. The identified anisotropic elasticity tensor of cortical bone from actual data based on the simplified model is given.
 [DOI: 10.1121/1.3087428]

PACS number(s): 43.40.At, 43.20.El, 43.20.Jr, 43.20.Gp [LLT] Pages: 2027–2034

I. INTRODUCTION

Elastic wave propagation can be used to characterize the anisotropic elasticity tensor of solids. For highly heterogeneous materials (e.g., concrete, wood, soft tissue, bone, etc.) not only the values of the stiffness constants may vary within the tested structure but also the material symmetry. For biological materials this is basically a consequence of their multiscale organization and of significant variations in the constitution at the nanoscopic scales. When the length scale of the heterogeneities is larger or of the order of magnitude of the wavelength, the determination of the elastic properties with acoustic waves is a challenge. If the acoustic signals used for material characterization have a large amount of scatter, the elastic tensor may be considered as not unique: the probed elastic properties depend on the actual wave path inside the material. In this context, the problem of elasticity identification may advantageously be formulated as the determination of a mean elasticity tensor and of the fluctuations of the elasticity around the mean tensor. In other words, since any model useful for the elasticity identification is always a rough approximation of the reality, the modeling of uncertainties may be considered in order to extend the domain of validity of the model.

The aim of this paper is to introduce a simplified elastoacoustic model useful for the ultrasonic characterization of a solid layer with uncertain elastic behavior. The determination of an uncertain elastic layer that represents a human compact bone is considered. However, the method developed is of a wider interest as it may be applied straightforward to the characterization of other materials. The complexity level of such a biomechanical system is such that a multiscale approach should be developed to represent all the mechanical behavior and would allow (among others) the ultrasonic characterization of a human compact bone to be simulated. Nevertheless, such a multiscale model would be very expensive and in this paper, we are only interested by the ultrasonic characterization of a human compact bone. So, such a multiscale model seems to be too "complete" for the present objectives. This is why a simplified model is introduced and is representative of *in vivo* measurements of bone properties with the so-called axial transmission technique (it should be noted that the other mechanical behaviors of the biomechanical system are not explored). Nevertheless, a question still remains. *In vivo* measurements exhibit a random fluctuation of the ultrasonic characterization which is not related to the

[a]Electronic mail: christophe.desceliers@univ-paris-est.fr

measurement errors related to the probe used. So it is investigated whether the proposed wave propagation model including a stochastic model of bone anisotropic elastic properties is able to account for *in vivo* measurements in a statistical sense. Thus, if such a stochastic simplified model has the same statistical properties as the *in vivo* measurements database, then it may be a good candidate for the detection of disease bone.

The axial transmission technique provides the velocity of ultrasonic waves axially transmitted along compact (cortical) bone through a linear arrangement of transmitters and receivers placed on the same side of the skeletal site. This technique has been extensively used to probe bone quality at the radius [see Bossy *et al.* (2004a) and Camus *et al.* (2000)]. In particular, the bidirectional axial transmission technique associated with a measurement of the velocity of the first arriving contribution of the signal (FAS) is a technique sensitive to the elasticity of radius cortical bone. Bossy *et al.* (2004b, 2004c), Grimal *et al.* (2007), and Raum *et al.* (2005) also suggested that bone mass density and bone thickness are other less important determinants of the FAS. Accordingly, as a matter of simplification, thickness will be considered as fixed in this work and density will be taken as a deterministic varying parameter of a probabilistic model to be identified as well as the mean elastic properties. For the purposes of the present work, experimental data previously collected with a bidirectional axial transmission device could be used to investigate the possible determination of compact bone stiffness tensor. In addition, a procedure to compensate the observed variations in soft tissue thickness under the probe is implemented in the probe. Accordingly, the possible variations in soft tissue thickness can be disregarded in the model.

The simplified elasto-acoustic model is a three-layer system: a solid layer sandwiched between two inviscid acoustic fluid layers. The elastic layer (cortical bone) is a homogeneous anisotropic elastic material while the fluid (acoustic) layers represent soft tissues. The uncertainties on the elasticity of bone are introduced via a probabilistic approach developed in Soize (2001, 2005). It should be noted that a first application of the parametric probabilistic approach to the axial transmission technique can be found in Macocco *et al.* (2006). The parameters of the probabilistic model are the mean value of the random elasticity tensor and an additional parameter that controls the statistical fluctuation level. The mean model for bone is transversely isotropic and the uncertain fluctuations are anisotropic.

The determination of the parameters of the model, including the parameters of the probabilistic model, is formulated as an optimization problem. Thus, the solution of this optimization problem yields an optimal probabilistic model for taking into account the uncertainties related to the elasticity tensor of the cortical layer. A family of probabilistic models for the cortical layer is introduced using information theory. Then, among all the probabilistic model of this family, the optimal one is identified. Hence, it should be noted that the mean values of the elasticity tensor are not expected to be accurately equal to those of the cortical bone of a given patient. Anyway, the experimental measurements used in this paper are not related to only one patient but to a set of several patients. So, pretending to identify the values of the actual elastic tensor for a given patient would be meaningless with the experimental database used. The algorithm used to calculate the values of the cost function for several test values of the parameters of the stochastic model is performed by means of a stochastic solver based on the Monte Carlo method associated with a hybrid numerical solver of wave propagation presented in Desceliers *et al.* (2008): for each independent realization of the random effective elasticity tensor, the transient elasto-acoustic response is calculated by the wave propagation solver.

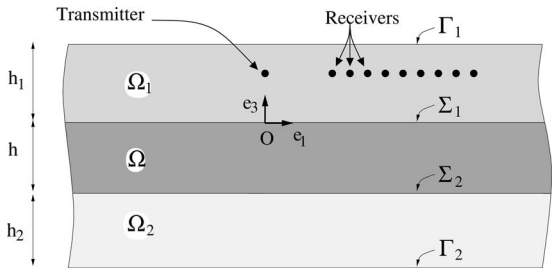


FIG. 1. Geometry of the multilayer system.

The capability of the stochastic model to represent the elasto-acoustic response of the uncertain system is demonstrated by a comparison of the model output to a database of *in vivo* measurements. In order to obtain a rich database from the statistical point of view, measurements from 168 patients were pooled; then the database takes into account a large variability of the properties of the elasto-acoustic system under consideration. Finally, the validation of the model consists in the comparison of the statistical properties of the model response to those of the *in vivo* measurements. Once validated, the model can be used to analyze the cortical bone stiffness tensor.

II. SIMPLIFIED MODEL OF THE MULTILAYER SYSTEM

The model consists of an anisotropic elastic homogeneous layer of thickness h sandwiched between two homogeneous inviscid acoustic fluids. The two fluids have identical density ρ_f and celerity c_f. The elastic layer is characterized by its mass density ρ and a stiffness tensor. In this work, the multilayer system is used to simulate experimental *in vivo* signals obtained with the so-called axial transmission technique.

The model geometry and the location of the acoustic source and receivers are shown in Fig. 1. The acoustic source is defined as a line source, placed inside Ω_1, and parallel to $\mathbf{e}_2$ where $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$ is an orthonormal basis for the space, $\mathbf{e}_3$ being normal to the fluid-solid interfaces. The wave field will be computed at each point $\mathbf{x}$ of the multilayer system with coordinate denoted (x_1,x_2,x_3) in the reference Cartesian frame $\mathbf{R}(O,\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$, where O is the origin placed on interface Σ_1.

The upper fluid layer (domain Ω_1) represents skin, muscle, and echographic gel used in the experiment to ensure a good contact between the ultrasonic probe and the skin. The elastic layer (domain Ω) represents the cortical bone while the lower fluid layer (Ω_2) models the marrow.

The thicknesses of the fluid layers Ω_1 and Ω_2 and of the elastic layer Ω are denoted h_1, h_2, and h, respectively. Free boundary conditions are applied on top of the upper fluid layer (surface Γ_1) and bottom of the lower fluid layer (surface Γ_2). In this paper, attention is focused on the biomechanical variability of bone elasticity. Accordingly, the density and the stiffness tensor of the elastic layer are considered as varying parameters of a probabilistic model to be identified while all the other parameters have fixed values. The varying parameters are identified following the procedure described in Sec. V.

III. PROBABILISTIC MODEL OF UNCERTAINTIES FOR THE ELASTICITY MATRIX OF THE SOLID LAYER

This section is devoted to the construction of a probabilistic model of the elastic material constituting the solid layer. The model accounts for uncertainties of the elastic constants in the bone material. That is, the bone elasticity is described by a mean stiffness tensor (with a prescribed symmetry: transverse isotropy in the paper) to which a random fluctuation is superimposed. Let c_{ijkh} be the components of the mean elasticity tensor on basis $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$. Then, mean elasticity matrix is defined as

$$[C] = \begin{pmatrix} [\tilde{C}_1] & [\tilde{C}_2] \\ [\tilde{C}_3] & [\tilde{C}_4] \end{pmatrix},$$

in which

$$[\tilde{C}_1] = \begin{pmatrix} c_{1111} & c_{1122} & c_{1133} \\ c_{2211} & c_{2222} & c_{2233} \\ c_{3311} & c_{3322} & c_{3333} \end{pmatrix},$$

$$[\tilde{C}_2] = \begin{pmatrix} \sqrt{2}c_{1123} & \sqrt{2}c_{1131} & \sqrt{2}c_{1112} \\ \sqrt{2}c_{2223} & \sqrt{2}c_{2231} & \sqrt{2}c_{2213} \\ \sqrt{2}c_{3323} & \sqrt{2}c_{3331} & \sqrt{2}c_{3312} \end{pmatrix},$$

$$[\tilde{C}_3] = \begin{pmatrix} \sqrt{2}c_{2311} & \sqrt{2}c_{2322} & \sqrt{2}c_{2333} \\ \sqrt{2}c_{3111} & \sqrt{2}c_{3122} & \sqrt{2}c_{3133} \\ \sqrt{2}c_{1211} & \sqrt{2}c_{1222} & \sqrt{2}c_{1233} \end{pmatrix},$$

$$[\tilde{C}_4] = \begin{pmatrix} 2c_{2323} & 2c_{2331} & 2c_{2312} \\ 2c_{3123} & 2c_{3131} & 2c_{3112} \\ 2c_{1223} & 2c_{1231} & 2c_{1212} \end{pmatrix}.$$

It should be noted that $[C_3]=[C_2]^T$. The construction of the probabilistic model consists in substituting $[C]$ by a random matrix $[\mathbf{C}]$ for which the probability density function is constructed using the information theory [see Shannon (1948, 1997)] with the available information defined as follows: (1) the random matrix $[\mathbf{C}]$ is a second-order random variable with values in the set $\mathbb{M}^+(\mathbb{R})$ of all the (6×6) real symmetric positive-definite matrices, (2) the mean value of random matrix $[\mathbf{C}]$ is the mean elasticity matrix $[C]$, and (3) the norm of the inverse matrix of $[\mathbf{C}]$ is a second-order random variable. It has been shown in Soize (2001, 2005) that the random matrix $[\mathbf{C}]$ can then be written as

$$[\mathbf{C}] = [L]^T[\mathbf{G}][L], \tag{1}$$

in which the (6×6) upper triangular matrix $[L]$ corresponds to the Cholesky factorization $[C]=[L]^T[L]$ and where the probability density function $p_{[\mathbf{G}]}$ of random matrix $[\mathbf{G}]$ is written as

$$p_{[\mathbf{G}]}([G]) = \mathbb{1}_{\mathbb{M}^+(\mathbb{R})}([G])c(\det[G])^b \exp\{-a\ \mathrm{tr}[G]\}, \tag{2}$$

where $a=7/(2\delta^2)$, $b=a(1-\delta^2)$, $\mathbb{1}_{\mathbb{M}^+(\mathbb{R})}([G])$ is equal to 1 if $[G]$ belongs to $\mathbb{M}^+(\mathbb{R})$ and is equal to zero if $[G]$ does not belong to $\mathbb{M}^+(\mathbb{R})$, $\mathrm{tr}[G]$ is the trace of matrix $[G]$ and where positive constant c is such that

$$c = \frac{(2\pi)^{-15/2}a^{6a}}{\prod_{j=1}^{6}\Gamma(\alpha_j)},$$

in which $\alpha_j=7/(2\delta^2)+(1-j)/2$ and where Γ is the gamma function. The parameter δ allows the dispersion of the random matrix $[\mathbf{C}]$ to be controlled. Thus, the parameters of the probabilistic model of uncertainties for the elasticity matrix are the components of $[C]$ and the coefficient δ. For such a probabilistic model, the random matrix $[\mathbf{G}]$ manifests the same statistical fluctuation in all its diagonal components and the same statistical fluctuation in all its extradiagonal components. It should be noted that the statistical fluctuations of the random matrix $[\mathbf{C}]$ are scaled with respect to the coefficients of the mean matrix $[C]$ [see Eq. (1)]. One of the important aspects of the probabilistic model is that it takes into account any anisotropic fluctuation of the elasticity tensor. However, the components of random matrix $[\mathbf{C}]$ are statistically dependent real-valued random variables. The generator of the independent realizations of random matrix $[\mathbf{G}]$ according to its probability density function [Eq. (2)] is given in the Appendix.

IV. COMPUTATION OF THE RANDOM TRANSIENT WAVE RESPONSE

A. Solving the equations

The transient elasto-acoustic response of the multilayer fluid-solid system is computed in the time-domain with a fast and efficient hybrid solver [see Desceliers *et al.* (2008)]. The symmetry of the problem allowed the previous problem to be reformulated in terms of the spatial coordinates x_1 and x_3 yielding a two-dimensional-space boundary value problem. The solver is based on a time-domain formulation associated with a one-dimensional-space Fourier transform for the infinite layer dimension (along the x_1 direction) and uses a finite element approximation in the direction perpendicular to the layers (along the x_3 direction). For a given mean elasticity matrix $[C]$, this solver allows the displacement field $\mathbf{u}$ in Ω and the pressure fields p_1 and p_2 in Ω_1 and Ω_2, respectively, to be calculated. Consequently, there exist three mappings g_{p_1}, $g_{\mathbf{u}}$ and g_{p_2} such that

$$p_1(\mathbf{x},t) = g_{p_1}(\mathbf{x},t;[C]), \tag{3}$$

$$\mathbf{u}(\mathbf{x},t) = g_{\mathbf{u}}(\mathbf{x},t;[C]), \tag{4}$$

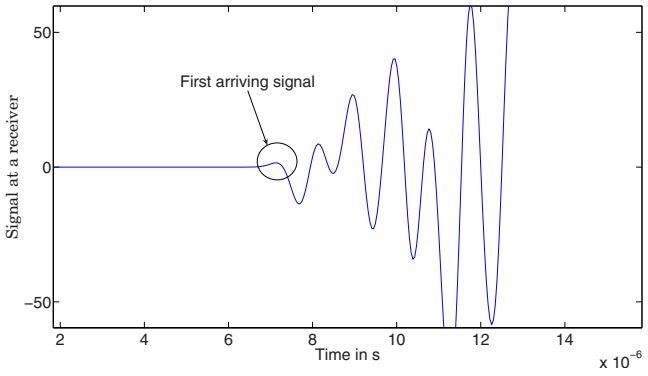


FIG. 2. (Color online) Example of a signal at receivers calculated by the model.

$$p_2(\mathbf{x},t) = g_{p_2}(\mathbf{x},t;[C]). \tag{5}$$

In addition, we introduce the von Mises stress at point $\mathbf{x} \in \Omega$ and at time t which is denoted by $\sigma^{\mathrm{vm}}(\mathbf{x},t)$. There also exists a mapping g_{vm} such that

$$\sigma^{\mathrm{vm}}(\mathbf{x},t) = g_{\mathrm{vm}}(\mathbf{x},t;[C]). \tag{6}$$

B. Velocity of the FAS and the axial transmission technique

The simulation of the actual measurements performed with the axial transmission technique consists in recording the signals $t \mapsto p_1(\mathbf{x},t)$ at several receivers located in Ω_1. The first arriving contribution of the signal is considered. Following the signal processing method used with the experimental device the velocity of FAS is determined from the time of flight of the first extremum of the contribution. Figure 2 shows a part of a simulated signal and the FAS. The simulated velocity v^{mod} of the FAS is estimated based on the different time arrivals at the consecutive receivers. Consequently, there exists a mapping g_{velo} such that

$$v^{\mathrm{mod}} = g_{\mathrm{velo}}(p_1). \tag{7}$$

C. Random transient wave response and random velocity of the FAS

The transient elasto-acoustic response of the stochastic multilayer fluid-solid system is random. Accordingly, the pressure and displacement fields are modeled by three random fields P_1, $\mathbf{U}$, and P_2 indexed by $\Omega_1 \times [0,T_{\max}]$, $\Omega \times [0,T_{\max}]$, and $\Omega_2 \times [0,T_{\max}]$ and with values in $\mathbb{R}$, $\mathbb{R}^3$, and $\mathbb{R}$, respectively. Substituting the mean elasticity matrix $[C]$ by the random matrix $[\mathbf{C}]$ in Eqs. (3)–(5) yields

$$P_1(\mathbf{x},t) = g_{p_1}(\mathbf{x},t;[\mathbf{C}]),$$

$$\mathbf{U}(\mathbf{x},t) = g_{\mathbf{u}}(\mathbf{x},t;[\mathbf{C}]),$$

$$P_2(\mathbf{x},t) = g_{p_2}(\mathbf{x},t;[\mathbf{C}]).$$

In addition, the velocity of the FAS and the von Mises stress are random variables denoted by V^{mod} and $\Sigma^{\mathrm{vm}}(\mathbf{x},t)$, respectively, such that [see Eqs. (6) and (7)]

$$\Sigma^{\mathrm{vm}}(\mathbf{x},t) = g_{\mathrm{vm}}(\mathbf{x},t;[\mathbf{C}]),$$

$$V^{\mathrm{mod}} = g_{\mathrm{velo}}(P_1). \tag{8}$$

V. IDENTIFICATION PROCEDURE OF THE STOCHASTIC MODEL

The objective of this section is to present a method to identify all the parameters of the stochastic model from a database of *in vivo* signals obtained previously. The parameters to identify are the components $[C]_{ij}$ of the mean elasticity matrix $[C]$, the coefficient δ that controls the statistical fluctuations of random elasticity tensor, and the mass density ρ.

A. Experimental database and random experimental velocity of the FAS

The signal used as input data for the determination of the parameters of the probabilistic model is associated with the wave propagation in the solid layer. This signal is referred to as FAS because, in practice, it always reaches the receivers first. This means that the other waves (direct wave from transmitter to receivers, waves guided by the fluid layer, etc.) arrive much later than the used signal. The physical reason for this is that the propagation path associated with the FAS includes a part in the solid for which the velocity is much higher than the fluid velocity. Accordingly, the FAS is not affected by the propagation in the top fluid layer and is representative of the solid plate acoustical properties. The latter was evidenced in simulations [see Bossy *et al.* (2004b), Camus *et al.* (2000), and Grimal and Naili (2006)] and experiments [see Bossy *et al.* (2004a)] which yield a FAS velocity of about 4000 m/s, which is the actual documented mean longitudinal wave velocity in compact bone.

In addition, the relative precision error on velocity measurements of the probe assessed on plates of academic materials such as Perspex or aluminum is lower than 0.2–0.3 [see, for instance, Bossy *et al.* (2004a)] which means that measurement errors can be neglected in the determination procedure.

The *in vivo* measurements were previously performed on a population of 168 subjects examined at the third distal radius. This group is a subset of a larger group of patients who participated to a clinical evaluation of the bidirectional axial transmission device. The multi-element probe operating at a center frequency of 1 MHz recorded 20 series of axially transmitted signals without particular angular scanning protocol except natural micro-movements of the operator. The experimental database finally consisted of 2018 measurements of FAS velocity. Each velocity measurement is considered as a realization of a random variable V^{exp} corresponding to the random variable V^{mod} obtained with the stochastic simplified model. The mean value of V^{exp} is $v^{\mathrm{exp}} = E\{V^{\mathrm{exp}}\}$ and its coefficient of variation Δ^{exp} is defined by $(\Delta^{\mathrm{exp}})^2 = E\{(V^{\mathrm{exp}})^2\}/(v^{\mathrm{exp}})^2 - 1$ in which $E\{\cdot\}$ is the mathematical expectation. Accordingly, the database consists of $N = 2018$ statistically independent realizations $V^{\mathrm{exp}}(\hat{\theta}_1), \ldots, V^{\mathrm{exp}}(\hat{\theta}_N)$ of random variable V^{exp}. Using the usual statistical

estimators and since N is sufficiently large, $v^{\exp}$ and $\Delta^{\exp}$ can reasonably be estimated by

$$v^{\exp} = \frac{1}{N}\sum_{k=1}^{N} V^{\exp}(\hat{\theta}_k),$$

$$\Delta^{\exp} = \frac{1}{v^{\exp}}\sqrt{\frac{1}{N}\sum_{k=1}^{N} V^{\exp}(\hat{\theta}_k)^2 - (v^{\exp})^2}.$$

B. Optimization problem for the identification procedure

For the identification procedure, it is assumed that the mean elasticity matrix represents a transverse isotropic homogeneous medium for which the plane (x_2,x_3) is the plane of isotropy. Then, all the components $[C]_{ij}$ are zeros except the following:

$$[C]_{11} = \frac{e_L^2(1-\nu_T)}{(e_L - e_L\nu_T - 2e_T\nu_L^2)},$$

$$[C]_{22} = \frac{e_T(e_L - e_T\nu_L^2)}{(1+\nu_T)(e_L - e_L\nu_T - 2e_T\nu_L^2)},$$

$$[C]_{12} = \frac{e_T e_L \nu_L}{(e_L - e_L\nu_T - 2e_T\nu_L^2)},$$

$$[C]_{23} = \frac{e_T(e_L\nu_T + e_T\nu_L^2)}{(1+\nu_T)(e_L - e_L\nu_T - 2e_T\nu_L^2)},$$

$$[C]_{44} = g_T,$$

$$[C]_{55} = g_L,$$

with $[C]_{22}=[C]_{33}$, $[C]_{12}=[C]_{13}=[C]_{21}=[C]_{31}$, $[C]_{23}=[C]_{32}$, and $[C]_{55}=[C]_{66}$ and where (1) e_L and e_T are the longitudinal and transverse Young moduli, (2) g_L and g_T are the longitudinal and transversal shear moduli, and (3) ν_L and ν_T are the longitudinal and transversal Poisson coefficients such that $g_T=e_T/2(1+\nu_T)$. Consequently, for this case, the matrix $[C]$ is written as $[C]=[\chi(e_L,\nu_L,g_L,e_T,\nu_T)]$ in which the function $[\chi]$ is completely defined. This equation shows that $[C]$ does not depend on 21 independent coefficients but only on 5 independent coefficients. Then, the unknown parameters of the stochastic model would be the coefficients e_L,ν_L,g_L,e_T,ν_T, the mass density ρ, and the coefficient δ. Let $\mathbf{a}$ be the vector such that $\mathbf{a}=(\rho,e_L,\nu_L,g_L,e_T,\nu_T)$. The problem consists in finding vector $\mathbf{a}$ and coefficient δ such that the stochastic model can represent the experimental database in a statistical sense. The optimal values $(\mathbf{a}^{\text{opt}},\delta^{\text{opt}})$ for $(\mathbf{a},\delta)$ are given by solving the following optimization problem:

$$(\mathbf{a}^{\text{opt}},\delta^{\text{opt}}) = \arg\min_{(\mathbf{a},\delta)} F^{\text{cost}}(\mathbf{a},\delta), \tag{9}$$

in which $F^{\text{cost}}(\mathbf{a},\delta)$ is a cost function which has to be defined. The optimization problem defined by Eq. (9) is solved by the simplex algorithm [see Nelder and Mead (1965)]. For each iteration of the simplex algorithm, the cost function has to be calculated which requires to solve the stochastic equations with Monte Carlo simulations (see Appendix).

1. Cost function associated with the usual least-square approach

The cost function $F^{\text{cost}}(\mathbf{a},\delta)$ associated with the usual least-square approach is defined as

$$F^{\text{cost}}(\mathbf{a},\delta) = E\{(V^{\exp} - V^{\text{mod}}(\mathbf{a},\delta))^2\},$$

in which $V^{\text{mod}}(\mathbf{a},\delta)$ is defined by Eq. (8) and where the dependence in $\mathbf{a}$ and δ is written. It should be noted that $V^{\exp}$ and $V^{\text{mod}}(\mathbf{a},\delta)$ are two independent random variables. Consequently, $F^{\text{cost}}(\mathbf{a},\delta)$ can be rewritten as

$$F^{\text{cost}}(\mathbf{a},\delta) = E\{(V^{\exp} - v^{\exp})^2\} + (v^{\exp} - v^{\text{mod}}(\mathbf{a},\delta))^2 + E\{(V^{\text{mod}}(\mathbf{a},\delta) - v^{\text{mod}}(\mathbf{a},\delta))^2\}, \tag{10}$$

in which $v^{\text{mod}}(\mathbf{a},\delta)=E\{V^{\text{mod}}(\mathbf{a},\delta)\}$. The first term in the right-hand side of Eq. (10) represents the variance of the random experimental velocity $V^{\exp}$ of the first arriving signal (which is independent of $\mathbf{a}$ and δ). The second term in the right-hand side of Eq. (10) corresponds to the bias between the stochastic simplified model and the experimental system. The last term in the right-hand side of Eq. (10) is variance of the random velocity V^{mod} of the first arriving signal calculated with the stochastic simplified model. It should be noted that this variance is equal to zero if $\delta=0$. Nevertheless, numerical experiments show that, for $\delta=0$, it is always possible to find a value $\mathbf{a}_0^{\text{opt}}$ of parameter $\mathbf{a}$ such that $v^{\exp} - v^{\text{mod}}(\mathbf{a},\delta)=0$. Consequently, the usual least-square approach yields $\mathbf{a}^{\text{opt}}=\mathbf{a}_0^{\text{opt}}$ and $\delta^{\text{opt}}=0$. Such a solution does not allow the statistical fluctuations due to uncertainties to be taken into account (because $\delta^{\text{opt}}=0$) which proves that the cost function defined by Eq. (10) is not adapted. This is the reason why a nonusual cost function is introduced below.

2. Cost function adapted to the problem

We introduce a cost function F^{cost} which is more sensitive to the statistical fluctuations of the model, i.e., to the values of δ. Such a cost function is written as

$$F^{\text{cost}}(\mathbf{a},\delta) = \frac{(v^{\exp} - v^{\text{mod}}(\mathbf{a},\delta))^2}{(v^{\exp})^2} + \frac{(\Delta^{\exp} - \Delta^{\text{mod}}(\mathbf{a},\delta))^2}{(\Delta^{\exp})^2},$$

in which

$$\Delta^{\text{mod}} = \sqrt{\frac{E\{V^{\text{mod}}(\mathbf{a},\delta)^2\}}{v^{\text{mod}}(\mathbf{a},\delta)^2} - 1}.$$

This cost function is used for the optimization problem defined by Eq. (9) in order to identify the parameters $\mathbf{a}$ and δ.

VI. RESULTS

A. Numerical data

The mean elasticity matrix of the cortical bone (layer Ω) is completely defined by the vector $\mathbf{a}=(\rho,e_L,\nu_L,g_L,e_T,\nu_T)$. The following numerical values for the components of $\mathbf{a}$ are used (1) for the direct simulations presented in Sec. VI B and (2) as initial parameters for the optimization presented in

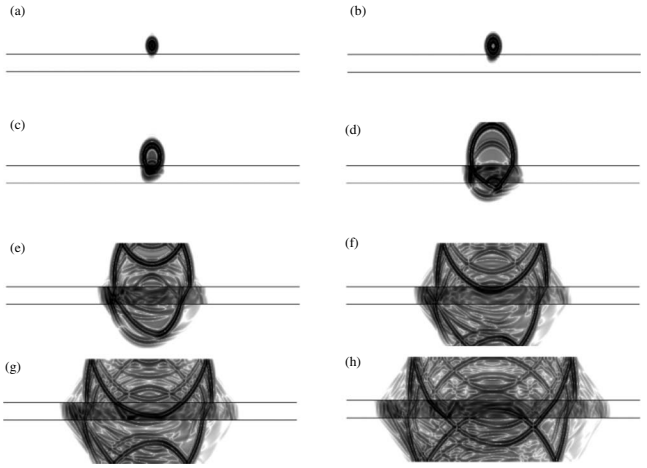


FIG. 3. Wave propagation in the three layers for a given realization of the random elasticity matrix at $t=1.56$ μs (a), $t=2.06$ μs (b), $t=2.94$ μs (c), $t=5.89$ μs (d), $t=9.72$ μs (e), $t=13.55$ μs (f), $t=15.31$ μs (g), and $t=119.88$ μs (h).

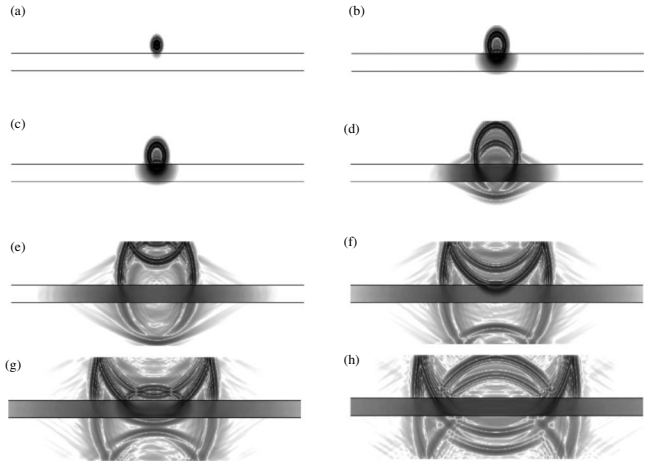


FIG. 4. Propagation of the mean waves in the three layers at $t=1.56$ μs (a), $t=2.06$ μs (b), $t=2.94$ μs (c), $t=5.89$ μs (d), $t=9.72$ μs (e), $t=13.55$ μs (f), $t=15.31$ μs (g), and $t=119.88$ μs (h).

Sec. VI C: $\rho=1722$ kg m^{-3}, $e_L=16.6$ GPa, $\nu_L=0.38$, $g_L=4.7$ GPa, $e_T=9.5$ GPa, $\nu_T=0.44$, $g_T=3.3$ GPa, $\rho_f=1000$ kg m^{-3}, and $c_f=1500$ m s^{-1}. The acoustic line source is located at $x_1^S=0$ and $x_3^S=2\times10^{-3}$ m and the time-history of the acoustic pulse is $F(t)=F_1\sin(2\pi f_c t)e^{-4(tf_c-1)^2}$, where $f_c=1$ MHz is the center frequency and $F_1=100$ N.

B. Analysis of the propagation of the uncertainties

In this section, we present an analysis of the propagation of the uncertainties in the multilayer system with $h_1=10^{-2}$ m, $h=4\times10^{-3}$ m, and $h_2=10^{-2}$ m. First, the wave propagation of one realization of the system is analyzed and clearly shows the effect of the anisotropy. Second, we analyze the mean value and the variance of the random wave propagation. This information allows the propagation of uncertainties to be quantified.

Let $[\mathbf{C}(\theta)]$ be one realization of the random elasticity matrix $[\mathbf{C}]$. For an elasticity matrix $[\mathbf{C}(\theta)]$, the realizations of the pressure fields in the two fluid media and the realization of the von Mises field in the solid medium are $P_1(\theta)$, $P_2(\theta)$, and $\Sigma^{\mathrm{vm}}(\theta)$ and such that

$$P_1(\mathbf{x},t;\theta)=g_{p_1}(\mathbf{x},t;[\mathbf{C}(\theta)]),$$

$$P_2(\mathbf{x},t;\theta)=g_{p_2}(\mathbf{x},t;[\mathbf{C}(\theta)]),$$

$$\Sigma^{\mathrm{vm}}(\mathbf{x},t;\theta)=g_{\mathrm{vm}}(\mathbf{x},t;[\mathbf{C}(\theta)]).$$

Figures 3(a)–3(h) show the graphs of the functions $(x_1,x_3)\mapsto P_1(x_1,x_3,t;\theta)$, $(x_1,x_3)\mapsto\Sigma^{\mathrm{vm}}(x_1,x_3,t;\theta)$, and $(x_1,x_3)\mapsto P_2(x_1,x_3,t;\theta)$ at different value of time t. Let $\underline{P}_1$, $\underline{\Sigma}^{\mathrm{vm}}$, and $\underline{P}_2$ be the mean values of the stochastic fields P_1, Σ^{vm}, and P_2. Figures 4(a)–4(h) show the graphs of functions $(x_1,x_3)\mapsto\underline{P}_1(x_1,x_3,t)$, $(x_1,x_3)\mapsto\underline{\Sigma}^{\mathrm{vm}}(x_1,x_3,t)$, and $(x_1,x_3)\mapsto\underline{P}_2(x_1,x_3,t)$ at different time t. Let var_{P_1}, $\mathrm{var}_{\Sigma^{\mathrm{vm}}}$, and var_{P_2} be the variances of the stochastic fields P_1, Σ^{vm}, and P_2. Figures 5(a)–5(h) show the graphs of the functions $(x_1,x_3)\mapsto\mathrm{var}_{P_1}(x_1,x_3,t)$, $(x_1,x_3)\mapsto\mathrm{var}_{\Sigma^{\mathrm{vm}}}(x_1,x_3,t)$, and $(x_1,x_3)\mapsto\mathrm{var}_{P_2}(x_1,x_3,t)$ at different time t.

Figures 3–5 taken together illustrate the stochastic wave propagation phenomenon. It can be seen in Fig. 3 that for one given realization, the wave field is particularly complex due to the anisotropy: in Fig. 3 (which corresponds to a realization $[\mathbf{C}(\theta)]$ of the stiffness matrix with a relative strong deviation from transverse anisotropy) the complex anisotropy is, in particular, evidenced by the asymmetry with respect to $(O,\mathbf{e}_3)$ of the wave field in the solid and in the bottom fluid layer. Such an asymmetry can be understood making use of slowness curves for anisotropic media [see, for instance, Royer and Dieulesaint (1999)]: if the slowness curves happen to be asymmetric with respect to $\mathbf{e}_3$, the waves at the right and left of $(O,\mathbf{e}_3)$ will travel at different velocities. In contrast, the mean and variance wave fields shown in Figs. 4 and 5 are quasisymmetric with respect to $(O,\mathbf{e}_3)$. This illustrates the fact that the anisotropic fluctuations of the stochastic elasticity matrix driven by δ do not have, on average, any preferential direction. Of course this can only be observed for the mean value and the variance. It should also be

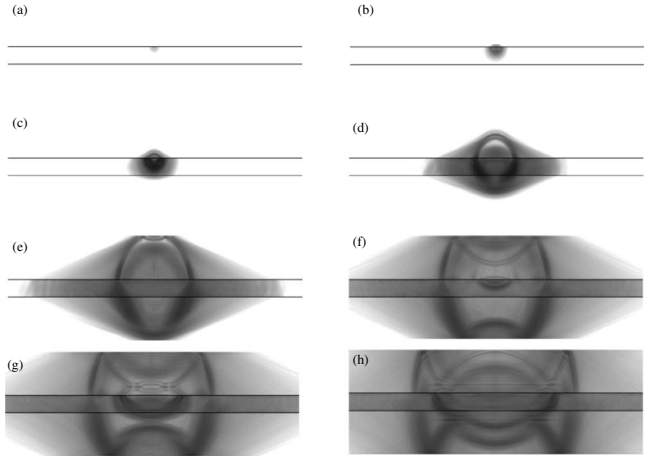


FIG. 5. Propagation of the uncertainties in the three layers at $t=1.56$ μs (a), $t=2.06$ μs (b), $t=2.94$ μs (c), $t=5.89$ μs (d), $t=9.72$ μs (e), $t=13.55$ μs (f), $t=15.31$ μs (g), and $t=119.88$ μs (h).

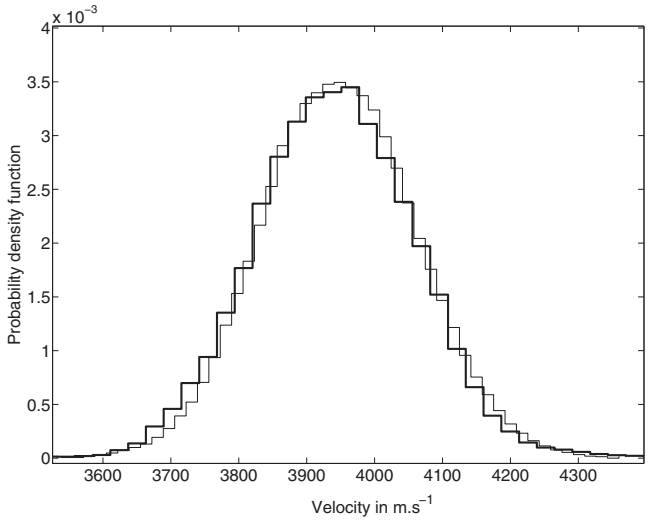


FIG. 6. Graph of the probability density function $v \mapsto p_{V^{\exp}}(v)$ (thick solid lines) and graph of the density probability density functions $v \mapsto p_{V^{\mathrm{mod}}}(v;\mathbf{a}^{\mathrm{opt}},\delta^{\mathrm{opt}})$ with $\mathbf{a}=\mathbf{a}^{\mathrm{opt}}$ and $\delta=\delta^{\mathrm{opt}}$ (thin solid lines).

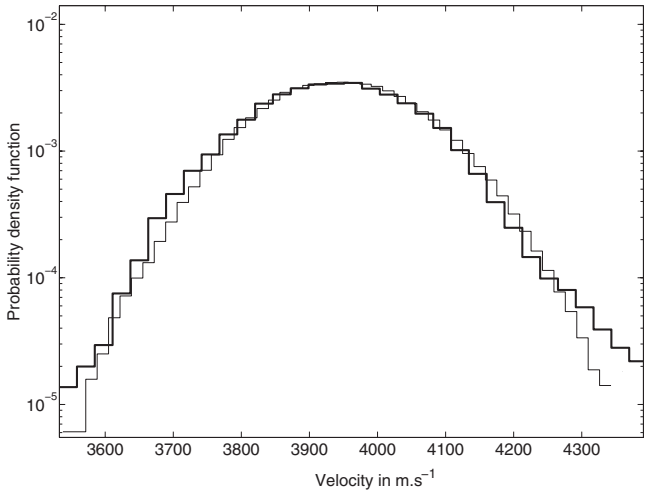


FIG. 7. Graphs of $v \mapsto \log(p_{V^{\exp}}(v))$ (thick solid lines) and $v \mapsto \log(p_{V^{\mathrm{mod}}}(v;\mathbf{a}^{\mathrm{opt}},\delta^{\mathrm{opt}}))$ (thin solid lines).

noted that the mean value of the different wave field realizations is not identical to the wave field obtained with the mean model unless $\delta=0$ (i.e., no statistical fluctuation of the elasticity tensor). Figure 5 shows that there is a significant propagation of the uncertainties whose intensity is relative to the variances. Note that the upper wavefront in Figs. 3 and 4 corresponding to the direct cylindrical wave from the source is missing in Fig. 5 while the other waves which have interacted with the solid layer are observed. This is an illustration of the fact that the uncertainties in the wave propagation problem are associated with an interaction with the random elastic layer. Furthermore the intensity of the uncertainties is higher where the amplitude of the wave (close to the wave front) is higher. In addition, it can be deduced from the comparison of Figs. 4 and 5 that the dispersion coefficient of the wave propagation is not the same everywhere in the multilayer system.

C. Experimental validation of the stochastic model and identification of the model parameters

The determination of the vector $\mathbf{a}=(\rho,e_L,\nu_L,g_L,e_T,\nu_T)$ and the coefficient δ was carried out using the method presented in Sec. V with $h_1=2\times10^{-3}$ m, $h=4\times10^{-3}$ m, and $h_2=10^{-2}$ m. The solutions $\mathbf{a}^{\mathrm{opt}}=(\rho^{\mathrm{opt}},e_L^{\mathrm{opt}},\nu_L^{\mathrm{opt}},g_L^{\mathrm{opt}},e_T^{\mathrm{opt}},\nu_T^{\mathrm{opt}})$ and δ^{opt} are such that $\rho^{\mathrm{opt}}=1598.8$ kg m^{-3}, $e_L^{\mathrm{opt}}=17.717$ GPa, $\nu_L^{\mathrm{opt}}=0.3816$, $g_L^{\mathrm{opt}}=4.7950$ GPa, $e_T^{\mathrm{opt}}=9.8254$ GPa, $\nu_T^{\mathrm{opt}}=0.4495$, and $\delta^{\mathrm{opt}}=0.1029$. For $\mathbf{a}=\mathbf{a}^{\mathrm{opt}}$ and $\delta=\delta^{\mathrm{opt}}$, the realizations $V^{\mathrm{mod}}(\mathbf{a},\delta;\theta_1),\ldots,V^{\mathrm{mod}}(\mathbf{a},\delta;\theta_N)$ of random velocity $V^{\mathrm{mod}}(\mathbf{a},\delta)$ are constructed with the stochastic simplified model and then, the probability density function $v \mapsto p_{V^{\mathrm{mod}}}(v;\mathbf{a},\delta)$ of $V^{\mathrm{mod}}(\mathbf{a},\delta)$ is estimated. Figure 6 shows in linear scale (1) the probability density function $v \mapsto p_{V^{\exp}}(v)$ of the random variable $V^{\exp}$ estimated with $N=2018$ experimental realizations $V^{\exp}(\hat{\theta}_1),\ldots,V^{\exp}(\hat{\theta}_N)$ and (2) the probability density function $v \mapsto p_{V^{\mathrm{mod}}}(v;\mathbf{a}^{\mathrm{opt}},\delta^{\mathrm{opt}})$. The same data are plotted in Fig. 7 in logarithmic scale in order to analyze the small levels of probabilities. These figures validate the stochastic simplified model in the statistical sense for the prediction of the velocity of the FAS.

VII. DISCUSSION AND CONCLUSIONS

A simplified elasto-acoustic multilayer model has been developed to simulate the ultrasonic wave propagation in a complex biomechanical system. In order to improve the simplified model, the uncertainties related to elastic behavior of the solid layer have been taken into account using a probabilistic approach. The determination of the stochastic model parameters was conducted for *in vivo* data collected previously with an ultrasonic axial transmission device.

Because the plots of the experimental and modeled probability density functions in Figs. 6 and 7 are so close, the stochastic simplified elasto-acoustic model is validated in the statistical sense regarding the prediction of the velocity of the FAS. Not only the mean value and the variance of the velocity are accurately predicted but also the experimental probability distribution is very well predicted; this remarkable result is not usual for such a complex system as that under consideration.

Furthermore, some values of the mean elastic constants of cortical bone and their fluctuations corresponding to the *in vivo* measurement database have been identified. As far as we know, the present work is the first attempt to identify the random anisotropic elasticity tensor of cortical bone from actual measurement data. Note that the elastic parameter of the mean model which, after the optimization, is the most changed compared to the provided nominal elasticity values is the Young modulus e_L in the axial direction. This is consistent with the physics of the wave propagation associated with the FAS when the thickness is large compared to the wavelength: basically the velocity of the observed FAS tends to the velocity of a head wave initiated on the flat fluid-solid interface [see Bossy *et al.* (2002)] and which is associated with bulk longitudinal wave propagation in the solid.

However, the identified elasticity values may not be compared straightforward to the actual bone elasticity values since they depend on the model configuration and on the fact that only the FAS contribution was considered. In the calcu-

lations, the thickness of the bone layer has been fixed among the largest physiological values; in these conditions, the FAS is known to be almost insensitive to the radius thickness [see Bossy *et al.* (2002)].

ACKNOWLEDGMENTS

This research was supported by the French Research National Agency (ANR) for the BONECHAR project.

APPENDIX: STOCHASTIC SOLVER

In this appendix, we present the stochastic solver that we have used to calculate the cost function F^{cost}. This stochastic solver is based on the Monte Carlo numerical method. First, the generator of independent realizations of the random matrix $[\mathbf{G}]$ presented in Soize (2001, 2005) is used. Random matrix $[\mathbf{G}]$ is written as $[\mathbf{G}]=[\mathbf{L}_G]^T[\mathbf{L}_G]$ in which $[\mathbf{L}_G]$ is a random upper triangular (6×6) real matrix whose random elements are independent variables defined as follows.

(1) For $j<j'$, the real-valued random variable $[\mathbf{L}_G]_{jj'}$ is written as $[\mathbf{L}_G]_{jj'}=\sigma U_{jj'}$ in which $\sigma=\delta/\sqrt{7}$ and where $U_{jj'}$ is a real-valued Gaussian random variable with zero mean and variance equal to 1.

(2) For $j=j'$, the positive-valued random variable $[\mathbf{L}_G]_{jj'}$ is written as $[\mathbf{L}_G]_{jj'}=\sigma\sqrt{2V_j}$ in which σ is defined above and where V_j is a positive-valued gamma random variable whose probability density function p_{V_j} with respect to dv is written as

$$p_{V_j}(v)=\mathbb{1}_{\mathbb{R}^+}(v)\frac{v^{(7/(2\delta^2)+(1-j)/2)}e^{-v}}{\Gamma(7/(2\delta^2)+(1-j)/2)}.$$

Then, n_s statistical independent realizations of random elasticity matrix $[\mathbf{C}]$ are constructed and are such that $[\mathbf{C}(\theta_1)]=[L]^T[\mathbf{G}(\theta_1)][L],\ldots,[\mathbf{C}(\theta_{n_s})]=[L]^T[\mathbf{G}(\theta_{n_s})][L]$. Let $P_1(\theta_1),\ldots,P_1(\theta_{n_s})$ be n_s statistical independent realizations of random field P_1 which are such that, for all $\mathbf{x}\in\Omega_1$ and $t\geq 0$, $P_1(\mathbf{x},t;\theta_1)=g_{p_1}(\mathbf{x},t;[\mathbf{C}(\theta_1)]),\ldots,P_1(\mathbf{x},t;\theta_{n_s})=g_{p_1}(\mathbf{x},t;[\mathbf{C}(\theta_{n_s})])$. As soon as the realizations of the random field P_1 are obtained, the realizations of the random velocity V^{mod} are deduced by $V^{\text{mod}}(\theta_1)=g_{\text{velo}}(P_1(\theta_1)),\ldots,V^{\text{mod}}(\theta_{n_s})=g_{\text{velo}}(P_1(\theta_{n_s}))$. For each value of $\mathbf{a}$ and δ, the value of the cost function $F^{\text{cost}}(\mathbf{a},\delta)$ is calculated by $F^{\text{cost}}_{n_s}$ (for sufficiently large values of n_s) such that

$$F^{\text{cost}}_{n_s}=\frac{(v^{\text{exp}}-v^{\text{mod}}_{n_s})^2}{(v^{\text{exp}})^2}+\frac{(\Delta^{\text{exp}}-\Delta^{\text{mod}}_{n_s})^2}{(\Delta^{\text{exp}}_N)^2},$$

in which $v^{\text{mod}}_{n_s}$ and $\Delta^{\text{mod}}_{n_s}$ are the estimations of v^{mod} and Δ^{mod} defined as

$$v^{\text{mod}}_{n_s}=\frac{1}{n_s}\sum_{k=1}^{n_s}V^{\text{mod}}(\theta_k),$$

$$\Delta^{\text{mod}}_{n_s}=\frac{1}{v^{\text{mod}}_{n_s}}\sqrt{\frac{1}{n_s}\sum_{k=1}^{n_s}V^{\text{mod}}(\hat{\theta}_k)^2-(v^{\text{mod}}_{n_s})^2}.$$

Bossy, E., Talmant, M., Defontaine, M., Patat, F., and Laugier, P. (**2004a**). "Bidirectional axial transmission can improve accuracy and precision of ultrasonic velocity measurement in cortical bone: A validation on test materials," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 71–79.

Bossy, E., Talmant, M., and Laugier, P. (**2002**). "Effect of bone cortical thickness on velocity measurements using ultrasonic axial transmission: A 2D simulation study," J. Acoust. Soc. Am. **112**, 297–307.

Bossy, E., Talmant, M., and Laugier, P. (**2004b**). "Three-dimensional simulations of ultrasonic axial transmission velocity measurement on cortical bone models," J. Acoust. Soc. Am. **115**, 2314–2324.

Bossy, E., Talmant, M., Peyrin, F., Akrout, L., Cloetens, P., and Laugier, P. (**2004c**). "An in vitro study of the ultrasonic axial transmission technique at the radius: 1-MHz velocity measurements are sensitive to both mineralization and intracortical porosity," J. Bone Miner. Res. **19**, 1548–1556.

Camus, E., Talmant, M., Berger, G., and Laugier, P. (**2000**). "Analysis of the axial transmission technique for the assessment of skeletal status," J. Acoust. Soc. Am. **108**, 3058–3065.

Desceliers, C., Soize, C., Grimal, Q., Haiat, G., and Naili, S. (**2008**). "A time-domain method to solve transient elastic wave propagation in a multilayer medium with a hybrid spectral-finite element space approximation," Wave Motion **45**(4), 383–399.

Grimal, Q., and Naili, S. (**2006**). "A theoretical analysis in the time-domain of wave reflection on a bone plate," J. Sound Vib. **298**, 12–29.

Grimal, Q., Raum, K., Desceliers, C., Soize, C., Haiat, G., Naili, S., Talmant, M., and Laugier, P. (**2007**). "Stochastic simulation of the axial transmission technique based on a multi-scale model of bone material properties," in *Second European Symposium on Ultrasonic Characterization of Bone*, Halle, Germany (Shaker, Aachen, Germany).

Macocco, K., Grimal, Q., Naili, S., and Soize, C. (**2006**). "Elastoacoustic model with uncertain mechanical properties for ultrasonic wave velocity prediction: Application to cortical bone evaluation," J. Acoust. Soc. Am. **119**, 729–740.

Nelder, J. A., and Mead, R. (**1965**). "A simplex-method for function minimization," Comput. J. **7**, 308–313.

Raum, K., Leguerney, I., Chandelier, F., Bossy, E., Talmant, M., Saied, A., Peyrin, F., and Laugier, P. (**2005**). "Bone microstructure and elastic tissue properties are reflected in QUS axial transmission measurements," Ultrasound Med. Biol. **31**, 1225–1235.

Royer, D., and Dieulesaint, E., *Elastic Waves in Solids: Free and Guided Propagation* (Springer, Berlin, **1999**), Vol. **1**.

Shannon, C. E. (**1948**). "A mathematical theory of communications," Bell Syst. Tech. J. **27**, 379–423.

Shannon, C. E. (**1997**). "A mathematical theory of communications (reprinted)," MD Comput. **14**, 306–317.

Soize, C. (**2001**). "Maximum entropy approach for modeling random uncertainties in transient elastodynamics," J. Acoust. Soc. Am. **109**, 1979–1996.

Soize, C. (**2005**). "Random matrix theory for modeling uncertainties in computational mechanics," Comput. Methods Appl. Mech. Eng. **194**, 1333–1366.

Weakly nonlinear oscillations of a compliant object buried in soil

Evgenia A. Zabolotskaya, Yurii A. Ilinskii, and Mark F. Hamilton
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029

(Received 5 August 2008; revised 4 December 2008; accepted 13 January 2009)

A nonlinear model equation in Rayleigh–Plesset form is developed for volume oscillations of a compliant object buried close to the surface in soil. The equation takes into account the stress-free boundary condition on the surface of the ground. The model is fully nonlinear given exact relations for the elastic potential energy stored in deformation of the object and the soil. Expansions of the potential energies for weak nonlinearity are provided in terms of elastic constants that can be determined experimentally. When the shear modulus is allowed to increase with depth below the surface, the natural frequency predicted by the model first decreases and thereafter increases with burial depth, in agreement with reported observations. Perturbation solutions are derived for the displacements on the surface of the ground at the second harmonic and difference frequency due to the nonlinear response of the object to acoustic excitation.
 [DOI: 10.1121/1.3079774]

PACS number(s): 43.40.Ga, 43.20.Tb, 43.25.Ts [JHG] Pages: 2035–2040

I. INTRODUCTION

Recent investigations[1–3] have revealed promising developments in the use of nonlinear acoustics for detection of land mines. Techniques based on the nonlinear response of a buried object to acoustic excitation can sometimes provide better discrimination between mines and other buried objects, and higher probability of detection, than obtained using linear methods.[3] It is thus desirable to have an appropriate model for the nonlinear response of a buried object to acoustical or seismic excitation. The model should account for the elastic properties of both the object and the surrounding soil.

The first such model was developed by Donskoy *et al.*[1] In this model the buried object and soil are characterized by lumped elements that obey nonlinear constitutive relations. The model predicts a continual decrease in the natural frequency with increasing burial depth. However, experiments have shown that the natural frequency ultimately increases with burial depth after passing through a minimum value.[2,4] The model was subsequently modified by Zagrai *et al.*[5] to account for soil stiffness that increases with burial depth in order to describe the eventual increase in natural frequency.

In a model developed by Sabatier and co-workers,[6,7] the soil is considered to be a continuous medium rather than a system of lumped elements. The continuous medium was modeled as an effective fluid, and thus elasticity of the soil is not taken into account. The natural frequency predicted by this model therefore decreases with depth until asymptotically approaching a limiting value. Nonlinear effects are also not taken into account. However, since the medium surrounding the object is continuous, compressible, and includes an interface corresponding to the surface of the ground, the model may be appropriate for scattering problems that cannot be investigated using a lumped-element model of the soil. The governing equations in the model must be solved numerically.

In the present paper, a nonlinear model equation is derived for the vibrational response of a compliant object buried in soil close to the surface and subjected to an applied acoustic pressure. Elasticity and nonlinearity of both the object and the soil are taken into account. The object is considered to be acoustically compact, and its dynamics are modeled as the radial pulsation of an elastic sphere. Since the frequencies of interest for mine detection are very low, typically several hundred hertz, and the objects are buried close to the surface of the ground, the region including the object and neighboring surface is in the reactive near field of the compressional waves radiated by or scattered from the object. The soil is therefore modeled as an incompressible elastic medium.

The equation of motion for such an object buried near the surface of an incompressible elastic half-space is derived using Lagrangian mechanics. Kinetic and potential energies of both the object and soil are taken into account. The result is an equation for the radial oscillation of the object that is expressed in Rayleigh–Plesset form. Correction terms that account for compressibility of the soil are shown to be very small. An expression is derived for the natural frequency in the linear approximation that accounts for both its initial decrease with depth due to increased inertia of the soil, and its eventual increase with depth due to increased rigidity of the soil. Perturbation solutions are derived for the displacement on the surface of the soil at the second harmonic and difference frequency resulting from a bifrequency acoustic pressure imposed on the object. Numerical calculations presented are based on material, soil, and acoustic source parameters reported in the literature.

II. THEORETICAL MODEL

We consider an acoustically compact object whose dynamical response to a disturbance may be characterized by its change in volume. The volume is in turn represented by

the instantaneous radius $R(t)$ of a sphere that occupies the same volume. The physical attributes of the object are characterized by a lumped inertia and stiffness. As justified at the end of this section, the continuous medium surrounding the object is presumed to be incompressible, with elasticity associated only with shear deformation. Lagrange's equation is used to derive the dynamical equation for the radial motion of the object:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{R}}\right) - \frac{\partial L}{\partial R} = 0 \tag{1}$$

where $L=T-U$ is the Lagrangian, T is the kinetic energy, and U is the potential energy of the system. Overdots on R and other quantities signify time derivatives.

Kinetic energy is stored in the motion of both the object and the surrounding soil:

$$T = \tfrac{1}{2}M\dot{R}^2 + 2\pi\rho R^3\dot{R}^2, \tag{2}$$

where M is the effective moving mass of the object, and ρ is the (constant) density of the soil. For example, M is the actual mass of a thin spherical shell, and it would be 0.6 times the mass of a solid elastic sphere. The second term in Eq. (2) is the kinetic energy in the soil.[8]

Potential energy in the system may be stored in deformation of both the object and the surrounding soil, and it may be altered due to work done by external forces. The expression for the potential energy is separated accordingly:

$$U = U_{\text{obj}} + U_{\text{soil}} + U_{\text{ext}}. \tag{3}$$

The potential energy stored in the object is expressed as a function of change in effective radius relative to the equilibrium value R_0:

$$U_{\text{obj}} = \frac{K_1}{2!}\left(\frac{R-R_0}{R_0}\right)^2 + \frac{K_2}{3!}\left(\frac{R-R_0}{R_0}\right)^3 + \cdots, \tag{4}$$

where the stiffness constants K_i have units of energy. The derivative required in Eq. (1) is

$$\frac{dU_{\text{obj}}}{dR} = \frac{1}{R_0}\left[K_1\left(\frac{R-R_0}{R_0}\right) + \frac{K_2}{2!}\left(\frac{R-R_0}{R_0}\right)^2 + \cdots\right], \tag{5}$$

which is the force required to displace the surface of the object from R_0 to R.

For comparison with experiment, the stiffness constants associated with the surface of the object may be more conveniently expressed as coefficients k_i having units of pressure per unit distance.[1] In this case, and in the linear approximation, Eq. (5) reduces to $dU_{\text{obj}}/dR = K_1R'/R_0^2$, where $R' = R-R_0$. Equating this force to $4\pi R_0^2 k_1 R'$ yields

$$K_1 = 4\pi R_0^4 k_1. \tag{6}$$

Similar relations are easily developed for higher-order coefficients. Frequently when considering nonlinearity in a spring, one ignores the quadratic (K_2) term in Eq. (5) and includes only the cubic (K_3) term (not shown). This practice is associated with the notion of a spring stiffness that increases as deviation from the equilibrium length increases, both in compression and tension. However, quadratic nonlinearity will be present in objects with complex geometries, and for sufficiently small deformations quadratic nonlinearity always dominates cubic nonlinearity.

The elastic energy stored by radial deformation in an incompressible medium may be expressed in terms of the expansion[8]

$$U_{\text{soil}} = 8\pi R_0^3\left[\mu\left(\frac{R-R_0}{R_0}\right)^2 - \frac{1}{2}\left(\mu+\frac{A}{3}\right)\left(\frac{R-R_0}{R_0}\right)^3 + \cdots\right], \tag{7}$$

where μ is the shear modulus and A is the third-order elastic constant in the notation of Landau and Lifshitz.[9] The required derivative is

$$\frac{dU_{\text{soil}}}{dR} = 16\pi R_0^2\left[\mu\left(\frac{R-R_0}{R_0}\right) - \frac{3}{4}\left(\mu+\frac{A}{3}\right)\left(\frac{R-R_0}{R_0}\right)^2 + \cdots\right], \tag{8}$$

which is the force required to displace the surface of the soil in contact with the object from R_0 to R.

Two external forces on the system are taken into account, and those are due to gravity and incident acoustic waves. Work done by these external forces changes the potential energy stored in the object (which is compressible) but not in the soil (which is incompressible), such that

$$U_{\text{ext}} = (\rho g d + p_{\text{ac}})(V - V_0), \tag{9}$$

where $\rho g d$ is the static pressure due to burial at depth d, g is the acceleration due to gravity, $p_{\text{ac}}(t)$ is the pressure perturbation due to an incident acoustic wave, $V=\frac{4}{3}\pi R^3$ is the volume of the object, and V_0 is its equilibrium value in the absence of static load. The required derivative is thus

$$\frac{dU_{\text{ext}}}{dR} = 4\pi R^2(\rho g d + p_{\text{ac}}). \tag{10}$$

Substitution of the expressions for T and U in Eq. (1) yields

$$\left(1+\frac{M}{4\pi\rho R^3}\right)R\ddot{R} + \frac{3}{2}\dot{R}^2 = -\frac{1}{4\pi\rho R^2}\left(\frac{dU_{\text{obj}}}{dR} + \frac{dU_{\text{soil}}}{dR}\right) - gd - \frac{p_{\text{ac}}(t)}{\rho}, \tag{11}$$

which is a modified form of the Rayleigh–Plesset equation for a bubble in an infinite liquid.

To complete the model we must satisfy the stress-free boundary condition on the surface of the ground. This condition can be satisfied by extending the medium throughout the half-space above the ground and introducing an image of the object that pulsates in opposite phase at height d above the plane where the surface had been located. At distance r from the object, the pressure variations in the free field due to volume oscillations of the object are described by $p = \rho\ddot{V}/4\pi r$. While this is the familiar expression for radiation from a monopole, it is not a linear approximation.[10] The pressure exerted on the object due to its image at distance $r=2d$ and with opposite phase will thus be $p_{\text{im}} = -\rho\ddot{V}/8\pi d$. Adding the term $-p_{\text{im}}/\rho = \ddot{V}/8\pi d$ to the right-hand side of Eq. (11) and using the relation $\ddot{V} = 4\pi R(R\ddot{R} + 2\dot{R}^2)$ yield

$$\left(1+\frac{M}{4\pi\rho R^3}-\frac{R}{2d}\right)R\ddot{R}+\left(\frac{3}{2}-\frac{R}{d}\right)\dot{R}^2 = -\frac{1}{4\pi\rho R^2}\left(\frac{dU_{\rm obj}}{dR}+\frac{dU_{\rm soil}}{dR}\right)-gd-\frac{p_{\rm ac}(t)}{\rho}. \quad (12)$$

The stress-free boundary condition is automatically taken into account by the dynamical equation for the radial oscillation of the object.

Equation (12) is the main result of this investigation. If the potential energies $U_{\rm obj}$ and $U_{\rm soil}$ are known, it is a fully nonlinear model for the forced pulsation of an elastic object buried near the surface of an elastic half-space. The remainder of the paper is devoted to approximate analytical solutions derived by perturbation under the condition of weak nonlinearity. Nonlinearity is included through quadratic order in the radial displacement, and therefore Eqs. (5) and (8) define all elastic constants required for calculations.

We conclude this section by assessing the effect of compressibility of the elastic medium. To do so, we return to Eq. (11) for an object in an infinite elastic medium. As shown elsewhere,[11] the effect of weak compressibility on a pulsating sphere in an infinite elastic medium is to impose a pressure (negative normal stress) $p_{\rm rad}=-\rho\ddot{V}/4\pi c_l$ on the sphere that is associated with radiation of the compressional wave at speed c_l. This pressure is taken into account by adding the term $-p_{\rm rad}/\rho=\ddot{V}/4\pi c_l$ to the right-hand side of Eq. (11). The stress-free boundary condition for an elastic half-space is then taken into account as above by adding the image term $\ddot{V}/8\pi d$ to the right-hand side. Compressibility introduces a finite time delay $2d/c_l$ for propagation of the compressional wave from the image to the object, such that the image term becomes $\ddot{V}(t-2d/c_l)/8\pi d$. Expansion of the image term in powers of $1/c_l$ yields $\ddot{V}(t)/8\pi d-\dddot{V}(t)/4\pi c_l$ to first order in $1/c_l$. The second term in the expansion thus cancels $-p_{\rm rad}/\rho$, indicating that for pulsation near a stress-free surface the correction for compressibility is of second order in $1/c_l$. The effect of compressibility in the context of the problem at hand is therefore very small. This conclusion is not surprising, because at low frequencies the pulsating object and its antiphase image form a dipole, the radiation efficiency of which is very poor at low frequencies.

III. LINEAR APPROXIMATION

Equation (12) is linearized with respect to the dimensionless radial displacement

$$\xi=(R-R_0)/R_0 \quad (13)$$

to obtain, taking Eqs. (5) and (8) into account,

$$\left(1+\frac{M}{4\pi\rho R_0^3}-\frac{R_0}{2d}\right)\ddot{\xi}+\left(\frac{K_1}{4\pi\rho R_0^5}+\frac{4\mu}{\rho R_0^2}\right)\xi = -\frac{gd}{R_0^2}-\frac{p_{\rm ac}(t)}{\rho R_0^2}. \quad (14)$$

In terms of the quantities

$$m=\frac{3M}{4\pi\rho R_0^3}, \quad \omega_{\rm obj}^2=\frac{K_1}{4\pi\rho R_0^5}, \quad \omega_{\rm soil}^2=\frac{4c_t^2}{R_0^2}, \quad (15)$$

where $c_t=(\mu/\rho)^{1/2}$ is the shear wave speed in the soil, Eq. (14) becomes

$$\left(1+\frac{m}{3}-\frac{R_0}{2d}\right)\ddot{\xi}+(\omega_{\rm obj}^2+\omega_{\rm soil}^2)\xi=-\frac{gd}{R_0^2}-\frac{p_{\rm ac}(t)}{\rho R_0^2}. \quad (16)$$

Alternatively, Eq. (16) may be expressed in the form

$$\ddot{\xi}+\omega_0^2\xi=-\frac{\omega_0^2}{\omega_{\rm obj}^2+\omega_{\rm soil}^2}\left(\frac{gd}{R_0^2}+\frac{p_{\rm ac}(t)}{\rho R_0^2}\right), \quad (17)$$

where the natural angular frequency ω_0 is defined by

$$\omega_0^2=\frac{\omega_{\rm obj}^2+\omega_{\rm soil}^2}{1+m/3-R_0/2d}, \quad (18)$$

$\omega_{\rm obj}$ is the natural frequency of an object with negligible mass in an infinite incompressible liquid, and $\omega_{\rm soil}$ is the natural frequency of a spherical cavity in an infinite incompressible elastic medium. The parameter m is the ratio of the effective moving mass of the object to the mass of the soil displaced by the presence of the object.

Due to the static load taken into account by the term $-gd/R_0^2$ in Eq. (16), free oscillations occur not about radius R_0 but instead about a local equilibrium radius $R_1<R_0$ determined by setting $\ddot{\xi}=0$ and $p_{\rm ac}=0$. The corresponding dimensionless displacement is

$$\xi_0=\frac{R_1-R_0}{R_0}=-\frac{gd}{R_0^2(\omega_{\rm obj}^2+\omega_{\rm soil}^2)}. \quad (19)$$

The magnitude of this static displacement may be estimated by considering a soft object ($\omega_{\rm obj}^2\ll\omega_{\rm soil}^2$), for which the expression reduces to $\xi_0=-gd/4c_t^2$. Taking $c_t=100$ m/s for soil and $d=1$ m for the burial depth one obtains $\xi_0=-2.5\times10^{-4}$. In the quadratic approximation, even for this excessively large burial depth, the static displacement is found to introduce a negligible change in the natural frequency, and it has no effect on generation of the second harmonic and difference frequency. The term $-gd$ in Eq. (12) is therefore ignored from this point forward.

As burial depth is increased, Eq. (18) assumes the limiting form

$$\omega_0^2=\frac{\omega_{\rm obj}^2+\omega_{\rm soil}^2}{1+m/3}, \quad d\gg R_0. \quad (20)$$

If the soil is homogeneous and therefore $\omega_{\rm soil}$ is a constant, then ω_0 is a monotonically decreasing function of depth, asymptotically approaching the lower bound determined by Eq. (20). This saturation effect was observed in experiments reported by Alberts *et al.*[7] on land mines immersed in water and buried in sand.

For numerical calculations we use equilibrium radius $R_0=0.1$ m and object stiffness $K_1=5\times10^5$ N m. The latter corresponds to the value $k_1=5\times10^8$ Pa/m from the table provided by Donskoy *et al.*,[1] and which is related to K_1 by Eq. (6). The density of the soil is taken to be $\rho=2000$ kg/m^3 and the damping coefficient to be $\delta=0.1$.

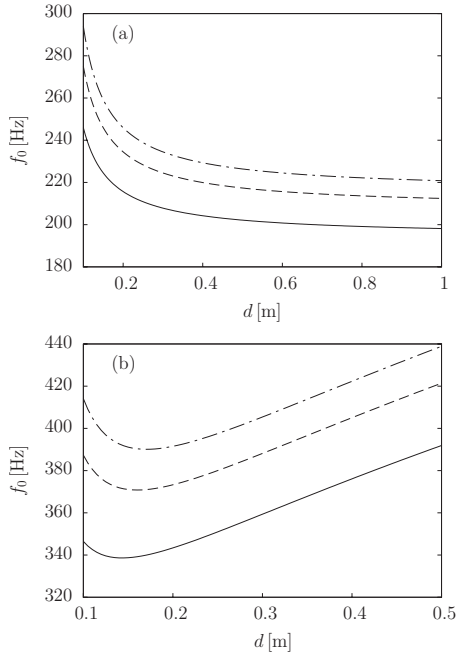


FIG. 1. Natural frequency as a function of burial depth for $R_0=0.1$ m, $\rho=2000$ kg/m^3, $K_1=5\times10^5$ N m, $\delta=0.1$, (a) $c_t=0$, and (b) $c_t=146d^{0.32}$ m/s, where d has units of meters, for $m=1$ (solid lines), $m=0.5$ (dashed lines), and $m=0.25$ (dot-dash lines).

Shown in Fig. 1(a) is the natural frequency $f_0=\omega_0/2\pi$ as a function of depth d for several values of the moving mass parameter m when the elasticity of the soil is ignored and therefore $\omega_{\text{soil}}=0$. It is observed that f_0 decreases as m increases, and that the asymptotic behavior described by Eq. (20) is reached at $d\sim1$ m. The frequency decreases because the pulsating object must move more soil when it is farther from the surface, which increases the effective mass loading on the object. As discussed above, the decrease in frequency associated with the increase in static load due to gravity is negligible.

Elasticity of the soil is now taken into account, and in particular its observed dependence on depth. Empirical relations developed by Nottis[12] for shear wave speeds as functions of depth, based on measurements made in various surficial materials along the lower Hudson Valley, permit calculation of natural frequency as a function of depth in inhomogeneous media. While evaluation of the integral resulting in Eq. (7) is based on the assumption that the coefficients μ, A, etc., are constant throughout the medium surrounding the object,[8] the effect of shear elasticity is in fact confined to a very small region in the neighborhood of the object. It is therefore reasonable to permit μ and thus c_t to vary locally with d in order to estimate the effect of inhomogeneity. On the basis of the relation obtained by Nottis[12] for glacial outwash sand and gravel, we let

$$c_t=100(d/0.305)^{0.32}=146d^{0.32}\ \text{m/s}, \tag{21}$$

where d in this equation possesses units of meters.

Taking all other parameters to be the same as in Fig. 1(a), one obtains the results in Fig. 1(b). Inhomogeneity characterized by a shear modulus that increases with depth thus eliminates saturation. Instead, after passing through a minimum value the natural frequency eventually increases with distance from the surface. Measurements of resonance frequency as a function of depth by Korman and Sabatier[2] and by Kang *et al.*[4] exhibit similar behavior.

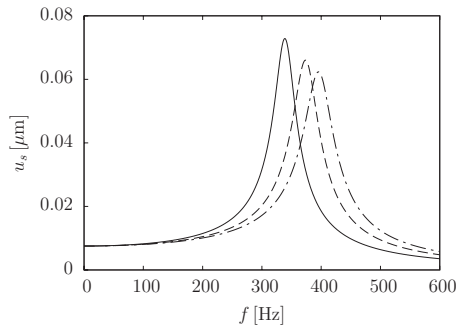


FIG. 2. Ground surface displacement at the drive frequency, as a function of drive frequency, for the same parameters as in Fig. 1(b) but with fixed depth $d=0.125$ m and acoustic pressure amplitude $p_0=5$ Pa, and again for $m=1$ (solid lines), $m=0.5$ (dashed lines), and $m=0.25$ (dot-dash lines).

Damping must be taken into account to model forced oscillations, and it is introduced *ad hoc* in Eq. (16) as follows:

$$\left(1+\frac{m}{3}-\frac{R_0}{2d}\right)(\ddot{\xi}+\delta\omega_0\dot{\xi})+(\omega_{\text{obj}}^2+\omega_{\text{soil}}^2)\xi=-\frac{p_{\text{ac}}(t)}{\rho R_0^2}. \tag{22}$$

The dimensionless damping coefficient δ is related to the quality factor Q by $\delta=1/Q$. It must be recognized that δ will inevitably vary with both depth and frequency due to the boundary condition imposed by the free surface and the composition of the object.

Forced excitation of the object is considered by representing the applied acoustic pressure as

$$p_{\text{ac}}=p_0\cos\omega t=\text{Re}[p_0e^{i\omega t}]. \tag{23}$$

The corresponding solution of Eq. (22) is expressed as

$$\xi(t)=\text{Re}[\Xi(\omega)e^{i\omega t}], \tag{24}$$

where the complex amplitude is

$$\Xi(\omega)=-\frac{p_0}{\rho R_0^2(\omega_{\text{obj}}^2+\omega_{\text{soil}}^2)(1-\omega^2/\omega_0^2+i\delta\omega/\omega_0)}. \tag{25}$$

Of particular interest is the displacement of the ground surface directly above the object. For a pulsating spherical source in an incompressible medium, the radial displacement u in the medium is inversely proportional to the square of the distance r from the center of the source, and thus $u=(R_0/r)^2\xi$. Directly above the object there will be doubling of the particle displacement at the free surface. For the magnitude of the ground surface displacement we thus write

$$u_s(\omega)=2(R_0/d)^2R_0|\Xi(\omega)|. \tag{26}$$

The surface displacement as a function of drive frequency $f=\omega/2\pi$ is plotted in Fig. 2 for the conditions in Fig. 1(b) but with fixed depth $d=0.125$ m and acoustic pressure amplitude $p_0=5$ Pa. The latter corresponds to a sound pressure level of approximately 105 dB (re 20 μPa) on the surface of the ground and is comparable to that used by Korman and

Sabatier,[2] as are the resulting surface displacements predicted in Fig. 2.

IV. QUADRATIC APPROXIMATION

In the quadratic approximation one obtains in place of Eq. (22)

$$\left(1+\frac{m}{3}-\frac{R_0}{2d}\right)(\ddot{\xi}+\delta\omega_0\dot{\xi})+(\omega_{\text{obj}}^2+\omega_{\text{soil}}^2)\xi = \Omega^2\xi^2-\left(1-\frac{2m}{3}-\frac{R_0}{d}\right)\xi\ddot{\xi}-\left(\frac{3}{2}-\frac{R_0}{d}\right)\dot{\xi}^2-\frac{p_{\text{ac}}(t)}{\rho R_0^2}, \tag{27}$$

where the quantity

$$\Omega^2=\left(2-\frac{K_2}{2K_1}\right)\omega_{\text{obj}}^2+\frac{1}{4}\left(11+\frac{A}{\mu}\right)\omega_{\text{soil}}^2 \tag{28}$$

contains the nonlinearity coefficients K_2 and A for the object and soil, respectively. The notation Ω^2 is employed to serve as a reminder in subsequent equations that the quantity has units of frequency squared.

A. Second harmonic

The acoustic excitation is given again by Eq. (23), for which the solution is now expressed as

$$\xi(t)=\text{Re}[\Xi_1(\omega)e^{i\omega t}+\Xi_2(\omega)e^{i2\omega t}]. \tag{29}$$

With $\Xi_1(\omega)$ and $\Xi_2(\omega)$ taken to be first- and second-order quantities, respectively, substitution of Eq. (29) in Eq. (27) yields

$$\Xi_2(\omega)=\frac{N_2(\omega)}{D_2(\omega)}\Xi_1^2(\omega) \tag{30}$$

at second order, where $\Xi_1(\omega)$ is given by Eq. (25), and with

$$N_2(\omega)=\Omega^2+\left(\frac{5}{2}-\frac{2m}{3}-\frac{2R_0}{d}\right)\omega^2, \tag{31}$$

$$D_2(\omega)=2(\omega_{\text{obj}}^2+\omega_{\text{soil}}^2)\left(1-\frac{4\omega^2}{\omega_0^2}+i\delta\frac{2\omega}{\omega_0}\right). \tag{32}$$

Figure 3(a) shows the surface displacement u_{2s} at the second harmonic as a function of source frequency for the same parameters as in Fig. 2. In addition, the values $K_2/K_1=10^5$ and $A/\mu=0$ were chosen for the nonlinearity parameters of the object and the soil, respectively. The motivation for choosing $A/\mu=0$ is that for soft elastic media such as loose soil $A/\mu=O(1)$,[13] in which case its contribution in Eq. (28) is negligible. We chose $K_2/K_1=10^5$ by determining the value required to predict surface displacements comparable to those measured at the second harmonic by Korman and Sabatier.[2] The frequency response observed in Fig. 3(a) is characteristic of nonlinear bubble oscillations, whereby the peak on the right results from driving the object at its natural frequency, and the peak on the left occurs when the second harmonic coincides with the natural frequency. In Fig. 3(b) the drive frequency is fixed at $f=340$ Hz while the depth is varied. The surface displacement drops rapidly as the burial depth is increased.

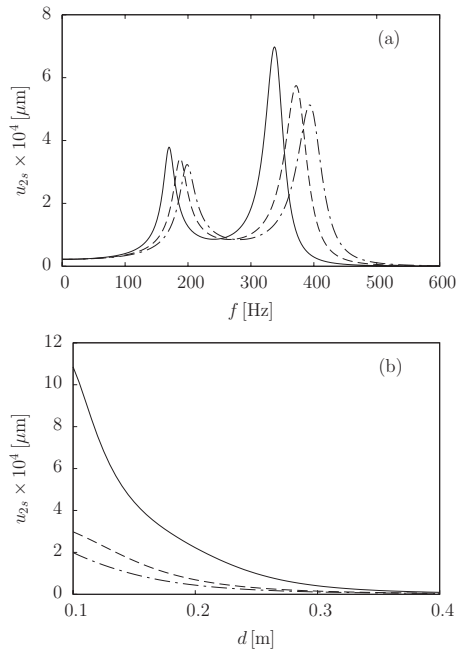


FIG. 3. Ground surface displacement at the second harmonic (a) as a function of drive frequency for fixed burial depth $d=0.125$ m, and (b) as a function of burial depth for fixed drive frequency $f=340$ Hz. The parameters are the same as in Fig. 2, but with $K_2/K_1=10^5$, $A/\mu=0$, and again for $m=1$ (solid lines), $m=0.5$ (dashed lines), and $m=0.25$ (dot-dash lines).

B. Difference frequency

We now let

$$p_{\text{ac}}=p_a\cos\omega_a t+p_b\cos\omega_b t=\text{Re}[p_a e^{i\omega_a t}+p_b e^{i\omega_b t}]. \tag{33}$$

The first-order response is expressed as

$$\xi_1(t)=\text{Re}[\Xi_{1a}(\omega_a)e^{i\omega_a t}+\Xi_{1b}(\omega_b)e^{i\omega_b t}], \tag{34}$$

where the complex amplitudes $\Xi_{1a}(\omega_a)$ and $\Xi_{1b}(\omega_b)$ are given by Eq. (25) with ω set equal to ω_a and ω_b, and p_0 set equal to p_a and p_b, respectively. The desired second-order response is now at the difference frequency $\omega_d=\omega_b-\omega_a$, with

$$\xi_d(t)=\text{Re}[\Xi_d(\omega_a,\omega_b)e^{i\omega_d t}]. \tag{35}$$

Equation (27) thus yields at second order

$$\Xi_d(\omega_a,\omega_b)=\frac{N_d(\omega_a,\omega_b)}{D_d(\omega_a,\omega_b)}\Xi_{1a}^*(\omega_a)\Xi_{1b}(\omega_b), \tag{36}$$

where the asterisk signifies complex conjugate, and

$$N_d(\omega_a,\omega_b)=2\Omega^2-\left(3-\frac{2R_0}{d}\right)\omega_a\omega_b+\left(1-\frac{2m}{3}-\frac{R_0}{d}\right)(\omega_a^2+\omega_b^2), \tag{37}$$

$$D_d(\omega_a,\omega_b)=2(\omega_{\text{obj}}^2+\omega_{\text{soil}}^2)\left(1-\frac{\omega_d^2}{\omega_0^2}+i\delta\frac{\omega_d}{\omega_0}\right). \tag{38}$$

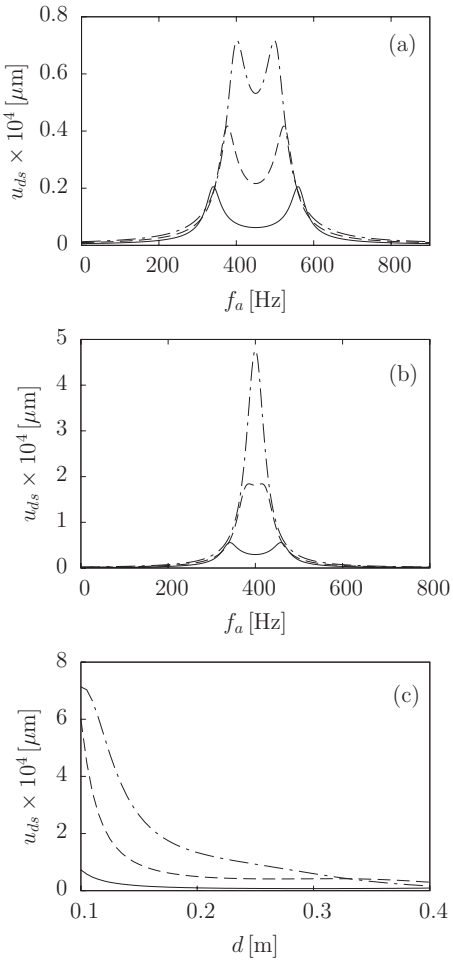


FIG. 4. Ground surface displacement at the difference frequency for the same material and geometric parameters as in Fig. 2(a), with $p_a=p_b=5$ Pa, and again for $m=1$ (solid lines), $m=0.5$ (dashed lines), and $m=0.25$ (dot-dash lines). (a) As a function of f_a for $f_b=900$ Hz. (b) As a function of f_a for $f_b=800$ Hz. (c) As a function of d for $f_a=400$ Hz and $f_b=800$ Hz.

Figure 4(a) shows the surface displacement u_{ds} at the difference frequency f_d as a function of the lower primary frequency f_a for fixed $f_b=900$ Hz. The parameters are otherwise the same as in Fig. 3(a), with $p_a=p_b=5$ Pa. The two maxima in each frequency response occur at the values of f_a for which $f_a=f_0$ and $f_b-f_a=f_0$, the center frequency of which is $f_a=f_b/2$. Figure 4(b) is based on the same parameters as in Fig. 4(a) except for $f_b=800$ Hz. Lowering the upper primary frequency f_b is seen to move the peaks closer together, and for $m=0.25$ only a single peak is observed. Figure 4(c) shows the dependence on burial depth for $f_a=400$ Hz and $f_b=800$ Hz.

V. CONCLUSION

The fully nonlinear model equation developed for the pulsation of a compliant object buried close to the surface in soil is expressed in Rayleigh–Plesset form. The stress-free condition on the surface of the ground is included in the model equation. The main restriction on the model is that the object size and burial depth are on the order of, or smaller than, the compressional wavelength in the medium. This condition permits the system dynamics to be determined completely by the volume oscillations of the object, and it allows the soil to be modeled as an incompressible elastic solid. In the linear approximation, the model predicts the natural frequency to first decrease and thereafter increase as burial depth is increased, in agreement with observations. The model yields simple analytical expressions for the response of the object and the displacement of the ground at the second harmonic and difference frequency of an acoustic excitation. The aforementioned features may make the model useful for investigating more complicated problems involving nonlinearity associated with scattering of elastic waves from a buried object.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research.

[1]D. Donskoy, A. Ekimov, N. Sedunov, and M. Tsionskiy, "Nonlinear seismo-acoustic land mine detection and discrimination," J. Acoust. Soc. Am. **111**, 2705–2714 (2002).
[2]M. S. Korman and J. M. Sabatier, "Nonlinear acoustic techniques for landmine detection," J. Acoust. Soc. Am. **116**, 3354–3369 (2004).
[3]D. Donskoy, A. Reznik, A. Zagrai, and A. Ekimov, "Nonlinear vibration of buried landmines," J. Acoust. Soc. Am. **117**, 690–700 (2005).
[4]W. Kang, J. A. Turner, F. Bobaru, L. Yang, and K. Rattanadit, "Granular layers on vibrating plates: Effective bending stiffness and particle-size effects," J. Acoust. Soc. Am. **121**, 888–896 (2007).
[5]A. Zagrai, D. Donskoi, and A. Ekimov, "Structural vibrations of buried land mines," J. Acoust. Soc. Am. **118**, 3619–3628 (2005).
[6]D. Velea, R. Waxler, and J. M. Sabatier, "An effective fluid model for landmine detection using acoustic to seismic coupling," J. Acoust. Soc. Am. **115**, 1993–2002 (2004).
[7]W. C. K. Alberts II, J. M. Sabatier, and R. Waxler, "Resonance frequency shift saturation in land mine burial simulation experiments," J. Acoust. Soc. Am. **120**, 1881–1886 (2006).
[8]S. Y. Emelianov, M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Nonlinear dynamics of a gas bubble in an incompressible elastic medium," J. Acoust. Soc. Am. **115**, 581–588 (2004).
[9]L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 3rd ed. (Pergamon, New York, 1986).
[10]Yu. A. Ilinskii and E. A. Zabolotskaya, "Cooperative radiation and scattering of acoustic waves by gas bubbles in liquid," J. Acoust. Soc. Am. **92**, 2837–2841 (1992).
[11]E. A. Zabolotskaya, Yu. A. Ilinskii, G. D. Meegan, and M. F. Hamilton, "Modifications of the equation for gas bubble dynamics in a soft elastic medium," J. Acoust. Soc. Am. **118**, 2173–2181 (2005).
[12]G. N. Nottis, "Predictive equations for soil shear-wave velocities: Lower Hudson Valley, New York," Multidisciplinary Center for Earthquake Engineering Research, Lamont-Doherty Earth Observatory, Columbia University (2001).
[13]M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Separation of compressibility and shear deformation in the elastic energy density," J. Acoust. Soc. Am. **116**, 41–44 (2004).

Elastodynamic wave scattering by finite-sized resonant scatterers at the surface of a horizontally layered halfspace

Geert Lombaert[a)]
Department of Civil Engineering, K.U.Leuven, Kasteelpark Arenberg 40, B-3001 Leuven, Belgium

Didier Clouteau
Laboratoire des Sols, Structures et Matériaux, Ecole Centrale de Paris, Grande Voie des Vignes, F-92295 Châtenay-Malabry Cedex, France

(Received 26 September 2008; revised 23 January 2009; accepted 30 January 2009)

The present paper deals with the multiple scattering by randomly distributed elastodynamic systems at the surface of a horizontally layered elastic halfspace due to an incident plane wave. Instead of solving this problem for a particular configuration of the system, multiple scattering theory is used to compute the ensemble response statistics. The Dyson equation is used to calculate the mean field, while the nonstationary second order statistics are obtained by means of the Bethe–Salpeter equation. This allows for the determination of the mean square response of the system in the time and frequency domains. This model is used to study multiple scattering between buildings under seismic excitation. The influence of multiple scattering on the seismic site response is verified. Furthermore, the influence of the footprint and the damping of the buildings are investigated. The results are compared to results of a coupled finite element/boundary element solution for a group of buildings. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3086290]

PACS number(s): 43.40.Ph, 43.20.Gp, 43.20.Px, 43.20.Tb [RLW] Pages: 2041–2052

I. INTRODUCTION

The dynamic behavior of a master structure coupled to complex substructures has received considerable attention in the past.[1–7] For vibroacoustic simulations of complex structures in the medium frequency range, Soize[1] introduced the concept of structural fuzzy. In this theory, the master structure is the main part of the system that is well represented by a deterministic mechanical model. The fuzzy substructure consists of a large number of secondary subsystems and is modeled by a probabilistic boundary impedance. Pierce *et al.*[3] and Strasberg and Feit[4] showed that for a large number of small substructures with closely spaced resonance frequencies, a damping of the master structure is obtained. This damping is due to a transfer of energy[2,5] from the primary structure to the secondary subsystems and therefore referred to as "apparent damping." Energy initially transferred to the subsystems can be returned to the master structure at later times. In the absence of true losses, the coupled system tends to an equipartition of energy.[2,5] Weaver[8,9] studied the energy flow for a plate in flexure interacting with randomly distributed sprung masses. The mean[8] and mean square response[9] are calculated based on multiple scattering theory.[10–13]

The problem of randomly distributed sprung mass systems interacting with an infinite plate in flexure is very similar to the one of the seismic response of urban sites. The buildings act as elastodynamic scatterers and can be considered as the locally reacting fuzzy substructure on top of the soil that plays the role of the wave-bearing master structure. Moreover, the interaction between the buildings is determined by resonant multiple scattering,[14] as the natural frequencies of high-rise buildings are in the frequency range of seismic excitation. This problem has been studied by several authors to more satisfactorily explain features of seismic records from urban arrays.[15] Wirgin and Bard[16] used a two-dimensional model of regularly spaced cylindrical blocks on top of a horizontally layered halfspace. Clouteau and Aubry[17] analyzed the response of a three-dimensional periodic city model, using a coupled finite element/boundary element method. Guéguen *et al.*[18] used an independent scattering approximation for a group of 180 buildings of the Roma Norte quarter of Mexico City. Mezher *et al.*[19] and Mezher[20] performed calculations for a group of 77 buildings of the same quarter, fully accounting for multiple wave scattering. Boutin and Roussillon[21] proposed a simplified analytical procedure, where the city is modeled as an equivalent layer at the soil's surface. Fernández-Ares and Bielak[22] used the domain reduction method[23] to study the seismic response of 88 sprung mass systems with a three-dimensional model taking into account the simulation of the earthquake source and propagation path effects, as well as the presence of the shallow basin underneath Mexico City. Kham *et al.*[24] studied periodically and nonperiodically spaced cities using two-dimensional boundary element models.

In the present paper, the model of randomly distributed elastodynamic systems at the surface of an elastic halfspace is used to study the seismic site response. The Dyson equation is used to calculate the mean response due to an incident plane wave, while the non-stationary second order statistics are obtained from the Bethe–Salpeter equation. Based on earlier developments,[25] a methodology is presented that takes into account the finite size of the interface between the scatterer and the halfspace. The results are complementary to those of Weaver[6,8] as the wave-bearing master structure is a

[a)]Electronic mail: geert.lombaert@bwk.kuleuven.be

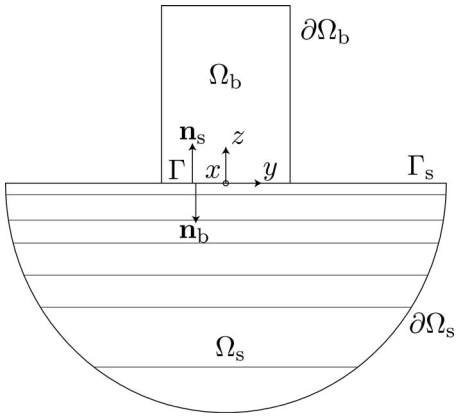


FIG. 1. The single scatterer problem.

horizontally layered elastic halfspace instead of a plate in flexure. Furthermore, the model is of large value for the interpretation of the results of the previously cited studies that consider the detailed response for a single city configuration with a limited number of buildings.

The outline of this paper is as follows. First, the equations of motion governing the single scatterer problem are formulated. An approximate expression is derived for the scattering operator in terms of the dynamic stiffness of the scatterer, the dynamic stiffness of the layered halfspace supporting a single scatterer, and the Green's functions of the halfspace. Next, the Foldy mean field approximation is considered. The mean Green's tensor of the coupled system and the mean response due to a vertically incident plane wave are calculated using Fourier transform techniques. The field correlation is obtained from the Foldy approximation of the Bethe–Salpeter equation. The configurationally averaged field quantities are used to study the seismic response of an urban site. The results are compared to a coupled finite element/boundary element solution for a group of 77 buildings.

II. THE SINGLE SCATTERER PROBLEM

A. Governing equations

In the following, the dynamic response of a single scatterer due to an incident wavefield is considered. The halfspace is assumed to be horizontally layered, and consists of a number of (homogeneous) layers with different Lamé coefficients λ and μ, and density ρ, on top of a homogeneous halfspace. The bounded elastic domain occupied by the scatterer is denoted as Ω_b, while Ω_s represents the unbounded elastic halfspace (Fig. 1). $\partial\Omega_b$ and $\partial\Omega_s$ are the boundaries of Ω_b and Ω_s, respectively. Γ_s is the part of $\partial\Omega_s$ corresponding to the free surface of the halfspace without the scatterer, while Γ is the common interface $\partial\Omega_b \cap \partial\Omega_s$. A right-handed Cartesian frame of reference $\mathbf{x}=(x,y,z)$ is used with its origin at the center of gravity of the interface Γ. For both subdomains, a linear elastic behavior is assumed, so that the displacement field $\mathbf{u}(\mathbf{x})$ satisfies the Navier equation in the frequency domain:

$$-\operatorname{div}\sigma(\mathbf{u})(\mathbf{x}) - \rho\omega^2\mathbf{u}(\mathbf{x}) = 0, \quad \mathbf{x} \in \Omega_b, \Omega_s, \tag{1}$$

where $\sigma(\mathbf{u})(\mathbf{x})$ denotes the elastic stress tensor at the point $\mathbf{x}$ for the displacement field $\mathbf{u}(\mathbf{x})$, ρ denotes the mass density, and ω denotes the circular frequency. It is assumed that no body forces are present, so that the right hand side of Eq. (1) is zero. The continuity of displacements and equilibrium of stresses at the common interface Γ are

$$\mathbf{u}_s(\mathbf{x}) = \mathbf{u}_b(\mathbf{x}), \quad \mathbf{x} \in \Gamma, \tag{2}$$

$$\sigma_s(\mathbf{u}_s)(\mathbf{x}) \cdot \mathbf{n}_s(\mathbf{x}) + \sigma_b(\mathbf{u}_b)(\mathbf{x}) \cdot \mathbf{n}_b(\mathbf{x}) = 0, \quad \mathbf{x} \in \Gamma, \tag{3}$$

where the subscript s refers to the elastic halfspace and the subscript b refers to the scatterer. For both subdomains, $\mathbf{n}(\mathbf{x})$ represents the outward unit normal vector to the boundary at point $\mathbf{x}$. At the common interface Γ, $\mathbf{n}_s(\mathbf{x}) = -\mathbf{n}_b(\mathbf{x})$. In the following, the tractions $\sigma_s(\mathbf{u}_s)(\mathbf{x}) \cdot \mathbf{n}_s(\mathbf{x})$ are denoted as $\mathbf{t}_s(\mathbf{u}_s)(\mathbf{x})$. An approximate solution of the problem is obtained, decomposing the displacements $\mathbf{u}_b(\mathbf{x})$ into the six rigid body modes $\boldsymbol{\psi}_m(\mathbf{x})$ and N structural modes $\boldsymbol{\varphi}_m(\mathbf{x})$ of the scatterer fixed on Γ:

$$\begin{aligned}\mathbf{u}_b(\mathbf{x}) &= \sum_{m=1}^{6} \boldsymbol{\psi}_m(\mathbf{x}) c_m + \sum_{m=1}^{N} \boldsymbol{\varphi}_m(\mathbf{x}) q_m \\ &= \Psi(\mathbf{x})\mathbf{c} + \Phi(\mathbf{x})\mathbf{q}, \quad \mathbf{x} \in \Omega_b,\end{aligned} \tag{4}$$

where c_m and q_m are the generalized coordinates. The rigid body modes are collected in the matrix $\Psi(\mathbf{x}) = [\boldsymbol{\psi}_1(\mathbf{x}) \ldots \boldsymbol{\psi}_6(\mathbf{x})]$, and the matrix $\Phi(\mathbf{x}) = [\boldsymbol{\varphi}_1(\mathbf{x}) \ldots \boldsymbol{\varphi}_N(\mathbf{x})]$ contains the structural modes. The generalized coordinates are collected in the vectors $\mathbf{c}$ and $\mathbf{q}$. In a cylindrical frame of reference (r,θ,z), the matrix $\Psi(\mathbf{x})$ is written as follows:

$$\Psi(r,\theta,z) = \begin{bmatrix} 1 & -z & 0 & 0 & 0 & -r\sin\theta \\ 0 & 1 & 0 & 0 & z & r\sin\theta \\ 0 & 0 & 1 & -r\sin\theta & r\cos\theta & 0 \end{bmatrix}, \tag{5}$$

where the center of gravity of Γ is the rotational center of the rotational rigid body modes.

In the halfspace, the displacement field $\mathbf{u}_s(\mathbf{x})$ is decomposed into the free field motion $\mathbf{u}_{\mathrm{ff}}(\mathbf{x})$, defined as the wavefield in the homogeneous layered halfspace Ω_s without the scatterer, and the diffracted wavefield $\mathbf{u}_d(\mathbf{x})$. The wavefield $\mathbf{u}_{\mathrm{ff}}(\mathbf{x})$ contains the incoming wave that propagates toward the free boundary $\partial\Omega_s$ as well as the waves that are reflected at the interfaces between layers. The diffracted wavefield $\mathbf{u}_d(\mathbf{x})$ represents the modification of the wavefield due to the scatterer and is decomposed into the locally diffracted wavefield $\mathbf{u}_{d0}(\mathbf{x})$ that cancels the free field motion $\mathbf{u}_{\mathrm{ff}}(\mathbf{x})$ at the interface Γ by definition, and the wavefields $\mathbf{u}_{dm}(\mathbf{x})$ radiated by the displacement modes $\psi_m(\mathbf{x})$:

$$\begin{aligned}\mathbf{u}_s(\mathbf{x}) &= \mathbf{u}_{\mathrm{ff}}(\mathbf{x}) + \mathbf{u}_{d0}(\mathbf{x}) + \sum_{m=1}^{6} \mathbf{u}_{dm}(\mathbf{x}) c_m \\ &= \mathbf{u}_{\mathrm{ff}}(\mathbf{x}) + \mathbf{u}_{d0}(\mathbf{x}) + \mathbf{U}_d(\mathbf{x})\mathbf{c}, \quad \mathbf{x} \in \Omega_s,\end{aligned} \tag{6}$$

where the matrix $\mathbf{U}_d(\mathbf{x}) = [\mathbf{u}_{d1}(\mathbf{x}) \cdots \mathbf{u}_{d6}(\mathbf{x})]$. At the interface Γ, $\mathbf{u}_{d0}(\mathbf{x}) = -\mathbf{u}_{\mathrm{ff}}(\mathbf{x})$ and $\mathbf{u}_{dm}(\mathbf{x}) = \boldsymbol{\psi}_m(\mathbf{x})$, so that the continuity

of displacements in Eq. (2) is satisfied. The equilibrium of stresses at the interface Γ in Eq. (3) is now elaborated as follows:

$$\mathbf{t}_s(\mathbf{u}_{\mathrm{ff}})(\mathbf{x}) + \mathbf{t}_s(\mathbf{u}_{d0})(\mathbf{x}) + \sum_{m=1}^{6} \mathbf{t}_s(\mathbf{u}_{dm})(\mathbf{x}) c_m + \sum_{m=1}^{6} \mathbf{t}_b(\boldsymbol{\psi}_m)(\mathbf{x}) c_m + \sum_{m=1}^{N} \mathbf{t}_b(\boldsymbol{\varphi}_m)(\mathbf{x}) q_m = 0. \tag{7}$$

A weak variational formulation of the equilibrium leads to the following system of equations:

$$\begin{bmatrix} \mathbf{K}^b_{qq} & \mathbf{K}^b_{qc} \\ \mathbf{K}^{bT}_{qc} & \mathbf{K}^b_{cc} + \mathbf{K}^s \end{bmatrix} \begin{pmatrix} \mathbf{q} \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{F}^s \end{pmatrix}, \tag{8}$$

where superscript T denotes the matrix transpose. The matrices $\mathbf{K}^b_{qq}$, $\mathbf{K}^b_{qc}$, and $\mathbf{K}^b_{cc}$ are calculated from the rigid body modes $\boldsymbol{\psi}_m(\mathbf{x})$ and the structural modes $\boldsymbol{\varphi}_m(\mathbf{x})$ of the scatterer fixed at Γ:

$$\mathbf{K}^b_{qq} = \mathrm{diag}(-\omega^2 + 2i\zeta_m\omega_m\omega + \omega_m^2), \tag{9}$$

$$\mathbf{K}^b_{qc} = -\rho_b\omega^2 \int_{\Omega_b} \Phi^T(\mathbf{x})\Psi(\mathbf{x})d\mathbf{x}, \tag{10}$$

$$\mathbf{K}^b_{cc} = -\rho_b\omega^2 \int_{\Omega_b} \Psi^T(\mathbf{x})\Psi(\mathbf{x})d\mathbf{x}, \tag{11}$$

where ω_m and ζ_m are the natural frequencies and damping ratios, respectively, corresponding to the structural mode $\boldsymbol{\varphi}_m$ and ρ_b is the mass density of the scatterer. In Eq. (8), $\mathbf{K}^s$ is the dynamic stiffness matrix of the halfspace. The elements $[\mathbf{K}^s]_{lm}$ of the matrix $\mathbf{K}^s$ are

$$[\mathbf{K}^s]_{lm} = \int_\Gamma \mathbf{t}_s(\mathbf{u}_{dm})(\mathbf{x}) \cdot \boldsymbol{\psi}_l(\mathbf{x})dS. \tag{12}$$

When the tractions $\mathbf{t}_s(\mathbf{u}_{dm})$ are collected in the matrix $\mathbf{T}_d = [\mathbf{t}_s(\mathbf{u}_{d1}) \cdots \mathbf{t}_s(\mathbf{u}_{d6})]$, Eq. (12) can be written as

$$\mathbf{K}^s = \int_\Gamma \mathbf{T}^T_d(\mathbf{x})\Psi(\mathbf{x})dS. \tag{13}$$

The force vector $\mathbf{F}^s$ in Eq. (8) represents the equivalent load due to the wavefield $\mathbf{u}_{\mathrm{ff}}(\mathbf{x})$. The elements $[\mathbf{F}^s]_l$ of the force vector are calculated from the tractions $\mathbf{t}_s(\mathbf{u}_{\mathrm{ff}})(\mathbf{x})$ of the wavefield $\mathbf{u}_{\mathrm{ff}}(\mathbf{x})$ and the tractions $\mathbf{t}_s(\mathbf{u}_{d0})(\mathbf{x})$ of the locally diffracted wavefield $\mathbf{u}_{d0}(\mathbf{x})$:

$$[\mathbf{F}^s]_l = -\int_\Gamma [\mathbf{t}_s(\mathbf{u}_{\mathrm{ff}})(\mathbf{x}) + \mathbf{t}_s(\mathbf{u}_{d0})(\mathbf{x})] \cdot \boldsymbol{\psi}_l(\mathbf{x})dS = -\int_\Gamma \mathbf{t}_s(\mathbf{u}_{\mathrm{ff}})(\mathbf{x}) \cdot \boldsymbol{\psi}_l(\mathbf{x})dS + \int_\Gamma \mathbf{t}_s(\mathbf{u}_{dl})(\mathbf{x}) \cdot \mathbf{u}_{\mathrm{ff}}(\mathbf{x})dS. \tag{14}$$

The second term has been elaborated, replacing $\boldsymbol{\psi}_l(\mathbf{x})$ by its equivalent $\mathbf{u}_{dl}(\mathbf{x})$ for $\mathbf{x} \in \Gamma$, using reciprocity, and making use of the fact that $\mathbf{u}_{d0}(\mathbf{x})$ is by definition equal to $-\mathbf{u}_{\mathrm{ff}}(\mathbf{x})$ for $\mathbf{x} \in \Gamma$. For a scatterer at the free surface of the halfspace, $\mathbf{t}_s(\mathbf{u}_{\mathrm{ff}})(\mathbf{x})$ vanishes for $\mathbf{x} \in \Gamma$:

$$\mathbf{F}^s = \int_\Gamma \mathbf{T}^T_d(\mathbf{x})\mathbf{u}_{\mathrm{ff}}(\mathbf{x})dS. \tag{15}$$

The calculation of the dynamic stiffness matrix $\mathbf{K}_s$ of the layered halfspace in Eq. (13) and the load $\mathbf{F}^s$ in Eq. (15) requires the solution of a boundary value problem where displacements $\mathbf{u}_{dm} = \boldsymbol{\psi}_m$ are prescribed at Γ and the corresponding tractions $\mathbf{t}_s(\mathbf{u}_{dm})$ are unknown. This problem is solved numerically by a boundary element method[26,27] based on regularized boundary integral equations.[28] For a scatterer at the free surface of a horizontally layered halfspace, the following boundary integral is derived for the tractions $\mathbf{t}_s(\mathbf{u}_{dm})$:

$$\psi_m(\mathbf{x}) = \int_\Gamma \mathbf{U}^G(\mathbf{x},\mathbf{x}')\mathbf{t}_s(\mathbf{u}_{dm})(\mathbf{x}')dS_{\mathbf{x}'} \quad \mathbf{x} \in \Gamma, \tag{16}$$

where $dS_{\mathbf{x}'}$ indicates that the surface integral is performed with respect to $\mathbf{x}'$, and $\mathbf{U}^G(\mathbf{x},\mathbf{x}')$ is the Green's tensor of the halfspace Ω_s. Each element $[\mathbf{U}^G]_{ij}(\mathbf{x},\mathbf{x}')$ represents the displacement at the position $\mathbf{x}'$ in the direction $\mathbf{e}_j$ due to a unit impulse load at the position $\mathbf{x}$ in the direction $\mathbf{e}_i$.[29,30] The Green's tensor $\mathbf{U}^G(\mathbf{x},\mathbf{x}')$ of a layered elastic halfspace[29,30] is obtained by means of the direct stiffness[31,32] or the reflection-transmission method.[33] In a boundary element method, Eq. (16) is solved by discretization of the interface Γ. Compared to other domain discretization techniques such as the finite element method, the advantage is that only Γ needs to be discretized and that energy radiation into the halfspace is modeled correctly through the use of Green's functions that satisfy Sommerfeld's radiation conditions.

The generalized coordinates $\mathbf{q}$ are now eliminated from Eq. (8):

$$(\mathbf{K}^b + \mathbf{K}^s)\mathbf{c} = \mathbf{F}^s, \tag{17}$$

where $\mathbf{K}^b$ is the dynamic stiffness matrix of the scatterer:

$$\mathbf{K}^b = \mathbf{K}^b_{cc} - \mathbf{K}^{bT}_{qc}(\mathbf{K}^b_{qq})^{-1}\mathbf{K}^b_{qc}. \tag{18}$$

For a scatterer at the free surface of a halfspace, the generalized coordinates $\mathbf{c}$ are found solving Eq. (17) with the generalized force (15) on the right hand side:

$$\mathbf{c} = (\mathbf{K}^b + \mathbf{K}^s)^{-1}\left[\int_\Gamma \mathbf{T}^T_d(\mathbf{x})\mathbf{u}_{\mathrm{ff}}(\mathbf{x})dS\right]. \tag{19}$$

B. The scattering operator for a single scatterer

The total wavefield $\mathbf{u}_s(\mathbf{x})$ in the layered halfspace is obtained by introducing solution (19) in Eq. (6). The diffracted wavefield $\mathbf{u}_d(\mathbf{x})$ now becomes

$$\mathbf{u}_d(\mathbf{x}) = \mathbf{u}_{d0}(\mathbf{x}) + \mathbf{U}_d(\mathbf{x})(\mathbf{K}^b + \mathbf{K}^s)^{-1}\left[\int_\Gamma \mathbf{T}_d^T(\mathbf{x})\mathbf{u}_{\mathrm{ff}}(\mathbf{x})dS\right]. \tag{20}$$

The wavefields $\mathbf{u}_{d0}$ and $\mathbf{u}_{dm}$ are computed from the corresponding stresses at Γ:

$$\begin{aligned}\mathbf{u}_d(\mathbf{x}) = &\int_\Gamma \mathbf{U}^G(\mathbf{x},\mathbf{x}')\mathbf{t}_s(\mathbf{u}_{d0})(\mathbf{x}')dS_{\mathbf{x}'} \\ &+ \left[\int_\Gamma \mathbf{U}^G(\mathbf{x},\mathbf{x}')\mathbf{T}_d(\mathbf{x}')dS_{\mathbf{x}'}\right] \\ &\times(\mathbf{K}^b + \mathbf{K}^s)^{-1}\left[\int_\Gamma \mathbf{T}_d^T(\mathbf{x}')\mathbf{u}_{\mathrm{ff}}(\mathbf{x}')dS_{\mathbf{x}'}\right].\end{aligned} \tag{21}$$

To further simplify this expression, $\mathbf{U}^G(\mathbf{x},\mathbf{x}')$ and the free field motion $\mathbf{u}_{\mathrm{ff}}(\mathbf{x})$ are approximated in terms of the modes $\Psi(\mathbf{x})$ for $\mathbf{x}\in\Gamma$:

$$\begin{aligned}\mathbf{u}_{\mathrm{ff}}(\mathbf{x}) &\approx \Psi(\mathbf{x})\left[\int_\Gamma \Psi^T(\mathbf{x})\Psi(\mathbf{x})dS\right]^{-1}\int_\Gamma \Psi^{\mathrm{T}}(\mathbf{x})\mathbf{u}_{\mathrm{ff}}(\mathbf{x})dS \\ &= \Psi(\mathbf{x})\left[\int_\Gamma \Pi^T(\mathbf{x})\mathbf{u}_{\mathrm{ff}}(\mathbf{x})dS\right], \quad \mathbf{x}\in\Gamma,\end{aligned} \tag{22}$$

$$\begin{aligned}\mathbf{U}^G(\mathbf{x},\mathbf{x}') &\approx \int_\Gamma \mathbf{U}^G(\mathbf{x},\mathbf{x}')\Psi(\mathbf{x}')dS_{\mathbf{x}'} \\ &\quad\times\left[\int_\Gamma \Psi^T(\mathbf{x})\Psi(\mathbf{x})dS\right]^{-1}\Psi^T(\mathbf{x}') \\ &= \left[\int_\Gamma \mathbf{U}^G(\mathbf{x},\mathbf{x}')\Pi(\mathbf{x}')dS_{\mathbf{x}'}\right]\Psi^{\mathrm{T}}(\mathbf{x}'), \quad \mathbf{x}'\in\Gamma,\end{aligned} \tag{23}$$

where the second equality in Eqs. (22) and (23) defines the matrix $\Pi(\mathbf{x})=\Psi(\mathbf{x})[\int_\Gamma \Psi^T(\mathbf{x})\Psi(\mathbf{x})dS]^{-1}$ that collects the scaled rigid body modes for any $\mathbf{x}\in\Gamma$. For scatterers that have a circular footprint with a radius R, Eq. (5) leads to the following expression for the matrix $\Pi(\mathbf{x})$ in a cylindrical frame of reference (r,θ,z):

$$\Pi(r,\theta) = \begin{bmatrix} \frac{1}{\pi R^2} & 0 & 0 & 0 & 0 & -\frac{2r\sin\theta}{\pi R^4} \\ 0 & \frac{1}{\pi R^2} & 0 & 0 & 0 & \frac{2r\sin\theta}{\pi R^4} \\ 0 & 0 & \frac{1}{\pi R^2} & -\frac{4r\sin\theta}{\pi R^4} & \frac{4r\cos\theta}{\pi R^4} & 0 \end{bmatrix}. \tag{24}$$

Introducing Eqs. (22) and (23) in Eq. (21) gives

$$\begin{aligned}\mathbf{u}_d(\mathbf{x}) = &-\left[\int_\Gamma \mathbf{U}^G(\mathbf{x},\mathbf{x}')\Pi(\mathbf{x}')dS_{\mathbf{x}'}\right] \\ &\times\int_\Gamma \mathbf{T}_d^T(\mathbf{x}')\mathbf{u}_{\mathrm{ff}}(\mathbf{x}')dS_{\mathbf{x}'} \\ &+\left[\int_\Gamma \mathbf{U}^G(\mathbf{x},\mathbf{x}')\Pi(\mathbf{x}')dS_{\mathbf{x}'}\right] \\ &\times\mathbf{K}^s(\mathbf{K}^b + \mathbf{K}^s)^{-1}\mathbf{K}^s \\ &\times\left[\int_\Gamma \Pi^T(\mathbf{x})\mathbf{u}_{\mathrm{ff}}(\mathbf{x})dS\right],\end{aligned} \tag{25}$$

where the first term is elaborated in a similar way as the force vector in Eq. (14) and Eq. (13) is used to rewrite the second term. Using approximation (22) for the wavefield $\mathbf{u}_{\mathrm{ff}}(\mathbf{x})$ in the first term allows combining both terms on the right hand side:

$$\begin{aligned}\mathbf{u}_d(\mathbf{x}) = &-\left[\int_\Gamma \mathbf{U}^G(\mathbf{x},\mathbf{x}')\Pi(\mathbf{x}')dS_{\mathbf{x}'}\right]\underbrace{[\mathbf{K}^s(\mathbf{K}^b + \mathbf{K}^s)^{-1}\mathbf{K}^b]}_{\mathbf{K}^{b'}} \\ &\times\left[\int_\Gamma \Pi^T(\mathbf{x}')\mathbf{u}_{\mathrm{ff}}(\mathbf{x}')dS_{\mathbf{x}'}\right]\end{aligned} \tag{26}$$

The scattered wavefield is obtained as the product of three terms. The first, rightmost term, is the excitation by the wavefield $\mathbf{u}_{\mathrm{ff}}$, obtained as its projection on the scaled rigid body modes $\Pi(\mathbf{x})$. The multiplication of the first and second terms gives the forces at the interface Γ. The second term is therefore defined as the modified dynamic stiffness $\mathbf{K}^{b'}$ of the scatterer. The third term redistributes the total forces at Γ and calculates the displacements in the halfspace Ω_s using the Green's tensor $\mathbf{U}^G(\mathbf{x}',\mathbf{x})$. The entire sequence of operations is the scattering operator. The dependence of $\mathbf{u}_d(\mathbf{x})$ on $\mathbf{x}$ is now rewritten as $\mathbf{u}_d(\mathbf{x},z)$, where $\mathbf{x}=(x,y)$. For a horizontally layered halfspace, the Green's tensor only depends on the relative horizontal position $\mathbf{x}'-\mathbf{x}$, and is written as $\mathbf{U}^G(\mathbf{x}'-\mathbf{x},z',z)$. For a scatterer at an arbitrary fixed position $(\mathbf{x}_b,z=0)$, Eq. (26) becomes

$$\begin{aligned}\mathbf{u}_d(\mathbf{x},z) = &-\left[\int_{\Gamma_s} \mathbf{U}^G(\mathbf{x}'-\mathbf{x},0,z)\Pi(\mathbf{x}'-\mathbf{x}_b)d\mathbf{x}'\right] \\ &\times\mathbf{K}^{b'}\left[\int_{\Gamma_s} \Pi^T(\mathbf{x}''-\mathbf{x}_b)\mathbf{u}_{\mathrm{ff}}(\mathbf{x}'',0)d\mathbf{x}''\right],\end{aligned} \tag{27}$$

where $\Pi^T(\mathbf{x}-\mathbf{x}_b)$ is zero for any point $\mathbf{x}\notin\Gamma$, so that the integrations over Γ can be replaced by integrations over the entire free surface $\Gamma_s=\mathbb{R}^2$ of the halfspace.

III. THE FOLDY MEAN FIELD APPROXIMATION

A. General formulation

In the following, the case is considered where a large number of scatterers are randomly distributed at the surface of the halfspace with an areal density ρ_{sc} and known scattering characteristics. The response of the coupled system due

to a vertically incident plane wave is studied. In order to solve this problem, a similar strategy as in the single scatterer case can be followed,[19,20] as will be discussed in Sec. VI. Instead, however, the configurationally averaged field quantities are calculated in an approximate way using multiple scattering theory. The approximation is based on truncated series expansions and only valid if the areal density of the scatterers is sufficiently small. Weaver used a similar methodology to calculate the mean response[8] and mean square response[9] for a plate interacting with randomly distributed sprung masses. More details regarding the derivation of the mean field approximations are found in Refs. 12 and 13 and more recent work by Weaver[8,9] and Photiades.[34]

Based on the solution (27) in the single scatterer case, the Foldy approximation[10,12,13,8] of the mean field $\underline{\mathbf{u}}(\mathbf{x})$ is formulated as follows:

$$\underline{\mathbf{u}}(\mathbf{x},z) = \mathbf{u}_{\mathrm{ff}}(\mathbf{x},z) - \rho_{sc}\int_{\Gamma_s} \mathbf{U}^G(\mathbf{x}'-\mathbf{x},0,z) \times \left\langle \int_{\Gamma_s} \Pi(\mathbf{x}'-\mathbf{x}_b)\mathbf{K}^{b'} \int_{\Gamma_s} \Pi^{\mathrm{T}}(\mathbf{x}''-\mathbf{x}_b) \right\rangle \times \underline{\mathbf{u}}(\mathbf{x}'',0)d\mathbf{x}''d\mathbf{x}_b d\mathbf{x}', \tag{28}$$

where the angular brackets denote the averaging with respect to the scatterer characteristics. A double Fourier transformation with respect to the horizontal coordinates $\mathbf{x}$ transforms the triple convolution into consecutive products in the wavenumber domain:

$$\tilde{\underline{\mathbf{u}}}(\mathbf{k},z) = \tilde{\mathbf{u}}_{\mathrm{ff}}(\mathbf{k},z) - \rho_{sc}\tilde{\mathbf{U}}^G(-\mathbf{k},0,z)\langle\tilde{\Pi}(\mathbf{k})\mathbf{K}^{b'}\tilde{\Pi}^T(-\mathbf{k})\rangle\tilde{\underline{\mathbf{u}}}(\mathbf{k},0), \tag{29}$$

where $\mathbf{k}$ collects k_x and k_y and the tilde denotes the representation of the variable in the wavenumber domain. An evaluation of this equation at $z=0$ allows for the calculation of the mean field $\tilde{\underline{\mathbf{u}}}(\mathbf{k},0)$ at the surface of the halfspace. The mean field $\tilde{\underline{\mathbf{u}}}(\mathbf{k},z)$ for any other z follows from a subsequent evaluation of the right hand side of Eq. (29).

When the scatterers have a circular footprint with a radius R, Eq. (24) can be used to compute the Hankel transform $\tilde{\Pi}(\mathbf{k})$ in the cylindrical wavenumber domain (k_r,ψ):

$$\tilde{\Pi}(\mathbf{k}) = \begin{bmatrix} \frac{2J_1(k_rR)}{k_rR} & 0 & 0 & 0 & 0 & -4i\sin\psi\frac{J_2(k_rR)}{k_rR^2} \\ 0 & \frac{2J_1(k_rR)}{k_rR} & 0 & 0 & 0 & 4i\cos\psi\frac{J_2(k_rR)}{k_rR^2} \\ 0 & 0 & \frac{2J_1(k_rR)}{k_rR} & -8i\sin\psi\frac{J_2(k_rR)}{k_rR^2} & 8i\cos\psi\frac{J_2(k_rR)}{k_rR^2} & 0 \end{bmatrix}, \tag{30}$$

where $J_1(x)$ and $J_2(x)$ are Bessel functions of the first kind, of orders 1 and 2, respectively. The relation between the Hankel transform in a cylindrical frame of reference and a double Fourier transform in a Cartesian frame[35] can be used to obtain the matrix $\tilde{\Pi}(\mathbf{k})$ in the wavenumber domain (k_x,k_y). In a similar way, an expression for the Green's tensor $\tilde{\mathbf{U}}^G(\mathbf{k},z',z)$ in the Cartesian wavenumber domain is obtained from its Hankel transform.[31]

The formulation of the ladder approximation of the Bethe–Salpeter equation in Sec. IV requires the mean Green's tensor $\tilde{\underline{\mathbf{U}}}^G(\mathbf{k},z,z')$ in the Foldy approximation:[9]

$$\tilde{\underline{\mathbf{U}}}^{GT}(\mathbf{k},z,z') = \tilde{\mathbf{U}}^{GT}(\mathbf{k},z,z') - \rho_{sc}\tilde{\mathbf{U}}^G(-\mathbf{k},0,z) \times\langle\tilde{\Pi}(\mathbf{k})\mathbf{K}^{b'}\tilde{\Pi}^T(-\mathbf{k})\rangle\tilde{\underline{\mathbf{U}}}^{GT}(\mathbf{k},0,z'). \tag{31}$$

For $z=0$, an equation in terms of $\tilde{\underline{\mathbf{U}}}^G(\mathbf{k},0,z')$ is obtained. The Green's tensor $\tilde{\underline{\mathbf{U}}}^G(\mathbf{k},z,z')$ for any other z is found from a subsequent evaluation of Eq. (31).

B. Foldy mean field for a vertically incident plane wave

When the response $\mathbf{u}_{\mathrm{ff}}(\mathbf{x},z)$ of the homogeneous layered halfspace Ω_s is generated by an incoming plane wave propagating through the underlying halfspace, the wavefield $\mathbf{u}_{\mathrm{ff}}(\mathbf{x},z)$ can be written in the frequency domain as

$$\mathbf{u}_{\mathrm{ff}}(\mathbf{x},z) = \mathbf{u}_{\mathrm{ff}}(z)\exp(-ik_y^s y)\pi\delta(\omega-\omega^s) + \mathbf{u}_{\mathrm{ff}}^*(z)\exp(ik_y^s y)\pi\delta(\omega+\omega^s), \tag{32}$$

where the superscript "*" denotes the complex conjugate, $\mathbf{u}_{\mathrm{ff}}(z)$ is the complex amplitude of the wavefield at depth z, and ω_s the circular frequency of the wave. The wavenumber k_y^s is found from the wavenumbers k_p and k_s for P- and S-waves in the underlying halfspace as $k_p\cos\alpha$ and $k_s\cos\alpha$, respectively, where α is the angle of incidence. The amplitude $\mathbf{u}_{\mathrm{ff}}(z)$ is computed from the amplitude of the incoming wave in the underlying halfspace by the direct stiffness method[31,32] or the reflection-transmission method.[33] The wavenumber domain representation $\tilde{\mathbf{u}}_{\mathrm{ff}}(\mathbf{k},z)$ is found by a double forward Fourier transform:

$$\tilde{\mathbf{u}}_{\mathrm{ff}}(\mathbf{k},z) = 4\pi^3 \mathbf{u}_{\mathrm{ff}}(z)\,\delta(k_x)\,\delta(k_y - k_y^s)\,\delta(\omega - \omega^s) + 4\pi^3 \mathbf{u}_{\mathrm{ff}}^*(z)\,\delta(k_x)\,\delta(k_y + k_y^s)\,\delta(\omega + \omega^s). \quad (33)$$

In the following, only the first term on the right hand side of this equation is considered. The contribution of the second term is found as the complex conjugate of the mean field due to the first term, with a frequency dependence $\delta(\omega+\omega^s)$ instead of $\delta(\omega-\omega^s)$. For an *SH*-wave with a polarization according to $\mathbf{e}_x$, the amplitude $\mathbf{u}_{\mathrm{ff}}(z)$ can be rewritten as $u_{\mathrm{ff}x}(z)\mathbf{e}_x$. Furthermore, the wave is assumed to be vertically incident, so that $\alpha=\pi/2$, $k_y^s=0$ and the delta function $\delta(k_y - k_y^s)$ reduces to $\delta(k_y)$. Based on Eq. (29), the Foldy approximation $\underline{\tilde{\mathbf{u}}}(\mathbf{k},z)$ of the mean field is obtained from the wavefield $\tilde{\mathbf{u}}_{\mathrm{ff}}(\mathbf{k},z)$ at the free surface $\Gamma_s(z=0)$ of the homogeneous layered halfspace Ω_s. Without loss of generality, it is assumed that the amplitude $\mathbf{u}_{\mathrm{ff}}(z)$ has a unit value at $z=0$. This leads to the following expression for $\underline{\tilde{\mathbf{u}}}(\mathbf{k},0)$:

$$\underline{\tilde{\mathbf{u}}}(\mathbf{k},0) = 4\pi^3 \delta(\mathbf{k})\,\delta(\omega - \omega^s)\mathbf{e}_x - \rho_{sc}\tilde{\mathbf{U}}^G(-\mathbf{k},0,0)\langle \tilde{\mathbf{\Pi}}(\mathbf{k})\mathbf{K}^{b\prime}\tilde{\mathbf{\Pi}}^T(-\mathbf{k})\rangle \underline{\tilde{\mathbf{u}}}(\mathbf{k},0), \quad (34)$$

where the product $\delta(k_x)\delta(k_y)$ is rewritten as $\delta(\mathbf{k})$. The mean field $\underline{\tilde{\mathbf{u}}}(\mathbf{k},0)$ can therefore be expressed as $4\pi^3\delta(\mathbf{k})\delta(\omega - \omega^s)\underline{\mathbf{u}}(0)$ with $\underline{\mathbf{u}}(0)$ the amplitude at $z=0$. Introducing this expression in Eq. (34) and performing an inverse Fourier transform from $\mathbf{k}$ to $\mathbf{x}$ give

$$\underline{\mathbf{u}}(\mathbf{x},0) = \mathbf{e}_x - \rho_{sc}\tilde{\mathbf{U}}^G(-\mathbf{k}=0,0,0) \times \langle \tilde{\mathbf{\Pi}}(\mathbf{k}=0)\mathbf{K}^{b\prime}\tilde{\mathbf{\Pi}}^T(-\mathbf{k}=0)\rangle \underline{\mathbf{u}}(\mathbf{x},0), \quad (35)$$

where the common factor $\pi\delta(\omega-\omega^s)$ has been dropped. For a limiting small wavenumber $\mathbf{k}\to 0$, the matrix $\tilde{\mathbf{\Pi}}(\mathbf{k})$ tends to

$$\lim_{\mathbf{k}\to 0} \tilde{\mathbf{\Pi}}(\mathbf{k}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \quad (36)$$

As $\tilde{\mathbf{U}}^G(-\mathbf{k}=0,0,0)$ is diagonal as well, the equations for different mean field components uncouple, and only the component $\underline{u}_x(0)$ corresponding to the polarization $\mathbf{e}_x$ of the incident wave is nonzero. For a given layered halfspace, the mean response $\underline{u}_x(0)$ only depends on the density ρ_{sc} and the first diagonal element $\langle[\mathbf{K}^{b\prime}]_{11}\rangle$ of the modified dynamic stiffness.

In the following, it is investigated how the distribution of scatterers at the surface of the halfspace affects the response of a single scatterer. The average response $\underline{\mathbf{u}}_b(\mathbf{x},0)$ at $z=0$ is calculated from the Foldy mean field $\underline{\mathbf{u}}(\mathbf{x},0)$:

$$\underline{\mathbf{u}}_b(\mathbf{x},0) = \left[\frac{1}{S_b}\int_\Gamma \Psi(\mathbf{x})dS\right]\langle(\mathbf{K}^b + \mathbf{K}^s)^{-1}\mathbf{K}^s\rangle \times \left[\int_\Gamma \Pi^T(\mathbf{x})\underline{\mathbf{u}}(\mathbf{x},0)dS\right], \quad (37)$$

where S_b is the area of the interface Γ and the first term between brackets is a matrix that contains the average displacement vector at Γ for each of the rigid body modes $\psi_m(\mathbf{x})$.

IV. THE BETHE–SALPETER EQUATION

A. General formulation of the ladder approximation

The Bethe–Salpeter equation[12] is an equation in terms of the correlation function $\mathbf{R}(\mathbf{x}',z',\mathbf{x}'',z'',t',t'')$ or, equivalently, $\langle \mathbf{u}(\mathbf{x}',z',t')\otimes\mathbf{u}(\mathbf{x}'',z'',t'')\rangle$, with $\otimes$ denoting the outer product. A representation in the frequency domain $\mathbf{S}(\mathbf{x}',z',\mathbf{x}'',z'',\omega',\omega'')$ is obtained by a double Fourier transform with respect to t' and t''. In the present case, the nonstationary Bethe–Salpeter equation in terms of $\mathbf{S}(\mathbf{x}',z',\mathbf{x}'',z'',\omega',\omega'')$ is formulated as follows:[12,36]

$$\begin{aligned}\mathbf{S}(\mathbf{x}',z',\mathbf{x}'',z'',\omega',\omega'') &= \underline{\mathbf{u}}(\mathbf{x}',z',\omega')\otimes\underline{\mathbf{u}}^*(\mathbf{x}'',z'',\omega'') \\ &+ \int_{\Gamma_s}\int_{\Gamma_s}\int_{\Gamma_s}\int_{\Gamma_s}[\underline{\mathbf{U}}^G(\mathbf{y}_1'-\mathbf{x}',0,z',\omega') \\ &\otimes \underline{\mathbf{U}}^{G*}(\mathbf{y}_1''-\mathbf{x}'',0,z'',\omega'')]:\mathbb{K}(\mathbf{y}_1',\mathbf{y}_1'',\mathbf{x}_1',\mathbf{x}_1'',\omega',\omega'') \\ &:\mathbf{S}(\mathbf{x}_1',0,\mathbf{x}_1'',0,\omega',\omega'')d\mathbf{y}_1'd\mathbf{y}_1''d\mathbf{x}_1'd\mathbf{x}_1'', \end{aligned} \quad (38)$$

where the fourth order tensor $\mathbb{K}(\mathbf{y}_1',\mathbf{y}_1'',\mathbf{x}_1',\mathbf{x}_1'',\omega',\omega'')$ is the kernel of the intensity operator, which is expressed as an infinite sum of operators using a diagram approach.[12,36] The operator ":" defines the double contraction of two tensors. For two fourth order tensors $\mathbb{A}$ and $\mathbb{B}$, the double contraction $\mathbb{A}:\mathbb{B}$ is a fourth order tensor $\mathbb{C}$ with $[\mathbb{C}]_{ijkl} = \Sigma_{m,n}[\mathbb{A}]_{ijmn}[\mathbb{B}]_{mnkl}$. The double contraction of the fourth order tensor $\mathbb{A}$ with the matrix $\mathbf{B}$ is the matrix $\mathbf{C}$, with $[\mathbf{C}]_{ij} = \Sigma_{k,l}[\mathbb{A}]_{ijkl}[\mathbf{B}]_{kl}$ when $\mathbf{C}=\mathbb{A}:\mathbf{B}$, and $[\mathbf{C}]_{kl}=\Sigma_{i,j}[\mathbf{B}]_{ij}[\mathbb{A}]_{ijkl}$ when $\mathbf{C}=\mathbf{B}:\mathbb{A}$. In the Foldy approach, the ladder approximation of the kernel of the intensity operator is[9]

$$\begin{aligned}\mathbb{K}(\mathbf{y}_1',\mathbf{y}_1'',\mathbf{x}_1',\mathbf{x}_1'',\omega',\omega'') &= \rho_{sc}\int_{\Gamma_s}\langle[\Pi(\mathbf{y}_1'-\mathbf{x}_b)\mathbf{K}^{b\prime}\Pi^T(\mathbf{x}_1'-\mathbf{x}_b)] \\ &\otimes[\Pi^*(\mathbf{y}_1''-\mathbf{x}_b)\mathbf{K}^{b\prime *}\Pi^\dagger(\mathbf{x}_1''-\mathbf{x}_b)]\rangle \\ &= \rho_{sc}\int_{\Gamma_s}[\Pi(\mathbf{y}_1'-\mathbf{x}_b)\otimes\Pi^*(\mathbf{y}_1''-\mathbf{x}_b)] \\ &:\langle\mathbf{K}^{b\prime}\otimes\mathbf{K}^{b\prime *}\rangle:[\Pi^{\mathrm{T}}(\mathbf{x}_1'-\mathbf{x}_b)\otimes\Pi^\dagger(\mathbf{x}_1''-\mathbf{x}_b)]d\mathbf{x}_b, \end{aligned} \quad (39)$$

where the superscript "†" denotes the Hermitian conjugate. In the derivation, the footprint of the scatterers and the matrix $\Pi(\mathbf{x})$ is assumed to be the same for all scatterers. As $\Pi(\mathbf{x})$ is real, $\Pi^*(\mathbf{y}_1''-\mathbf{x}_b)=\Pi(\mathbf{y}_1''-\mathbf{x}_b)$ and $\Pi^\dagger(\mathbf{x}_1''-\mathbf{x}_b)=\Pi^{\mathrm{T}}(\mathbf{x}_1''-\mathbf{x}_b)$ in Eq. (39).

B. Foldy field correlation approximation for a vertically incident plane wave

For a vertically incident plane wave, the correlation $\mathbf{S}(\mathbf{x}',z',\mathbf{x}'',z'')$ only depends on the relative horizontal position $\mathbf{x}''-\mathbf{x}'$ and is written as $\mathbf{S}(\mathbf{x}''-\mathbf{x}',z'',z')$. In a similar way as for the mean field, the solution $\mathbf{S}(\mathbf{x}''-\mathbf{x}',0,0)$ for $z''=z'=0$ allows for the evaluation of $\mathbf{S}(\mathbf{x}''-\mathbf{x}',z'',z')$ for any

(z'', z'). The vertical coordinates are therefore omitted, assuming an implicit evaluation at $z''=0$ and $z'=0$. This leads to the following expression:

$$\mathbf{S}(\mathbf{x}''-\mathbf{x}') = \underline{\mathbf{u}}(\mathbf{x}') \otimes \underline{\mathbf{u}}^*(\mathbf{x}'') + \rho_{sc}\int_{\Gamma_s}\int_{\Gamma_s}\int_{\Gamma_s} [\underline{\mathbf{U}}^G(\mathbf{y}_1'-\mathbf{x}') \times \Pi(\mathbf{y}_1'-\mathbf{x}_b) \otimes \underline{\mathbf{U}}^{G*}(\mathbf{y}_1'-\mathbf{x}'')\Pi(\mathbf{y}_1'-\mathbf{x}_b)] :\langle \mathbf{K}^{b'} \otimes \mathbf{K}^{b'*}\rangle : \left[\int_{\Gamma_s}\int_{\Gamma_s} \Pi^T(\mathbf{x}_1'-\mathbf{x}_b) \otimes \Pi^T(\mathbf{x}_1'' - \mathbf{x}_b) : \mathbf{S}(\mathbf{x}_1''-\mathbf{x}_1')d\mathbf{x}_1' d\mathbf{x}_1''\right] d\mathbf{x}_b d\mathbf{x}_1' d\mathbf{x}_1'', \quad (40)$$

where the dependence on ω' and ω'' is understood implicitly as well. The final bracketed term is defined as the matrix $\mathbf{S}_\Pi(\mathbf{x}''-\mathbf{x}')$, evaluated at $\mathbf{x}''=\mathbf{x}'=\mathbf{x}_b$:

$$\mathbf{S}_\Pi(\mathbf{x}''-\mathbf{x}') = \int_{\Gamma_s}\int_{\Gamma_s} [\Pi^{\mathrm{T}}(\mathbf{x}_1'-\mathbf{x}') \otimes \Pi^{\mathrm{T}}(\mathbf{x}_1''-\mathbf{x}'')] : \mathbf{S}(\mathbf{x}_1''-\mathbf{x}_1')d\mathbf{x}_1' d\mathbf{x}_1''. \quad (41)$$

Equation (40) is formulated in the wavenumber domain by a Fourier transform with respect to $\mathbf{x}''-\mathbf{x}'$:

$$\tilde{\mathbf{S}}(\mathbf{k}) = \tilde{\underline{\mathbf{S}}}(\mathbf{k}) + \rho_{sc}[\tilde{\underline{\mathbf{U}}}^G(\mathbf{k})\tilde{\Pi}(-\mathbf{k}) \otimes \tilde{\underline{\mathbf{U}}}^{G*}(\mathbf{k})\tilde{\Pi}(\mathbf{k})] :\langle \mathbf{K}^{b'} \otimes \mathbf{K}^{b'*}\rangle : \mathbf{S}_\Pi(0), \quad (42)$$

where $\tilde{\underline{\mathbf{S}}}(\mathbf{k})$ is the Fourier transform of $\underline{\mathbf{u}}(\mathbf{x}') \otimes \underline{\mathbf{u}}^*(\mathbf{x}'')$. Equation (42) is transformed into an equation in terms of $\mathbf{S}_\Pi(0)$, applying the operator $[\tilde{\Pi}^T(\mathbf{k}) \otimes \tilde{\Pi}^T(-\mathbf{k})]$: on both the left and right hand sides, integrating over all wavenumbers and multiplying by $1/4\pi^2$:

$$\mathbf{S}_\Pi(0) = \underline{\mathbf{S}}_\Pi(0) + \rho_{sc}\left[\frac{1}{4\pi^2}\int_{\mathbb{R}^2} \tilde{\Pi}^T(\mathbf{k})\tilde{\underline{\mathbf{U}}}^G(\mathbf{k})\tilde{\Pi}(-\mathbf{k}) \otimes \tilde{\Pi}^T(-\mathbf{k})\tilde{\underline{\mathbf{U}}}^{G*}(\mathbf{k})\tilde{\Pi}(\mathbf{k})d\mathbf{k}\right] :\langle \mathbf{K}^{b'} \otimes \mathbf{K}^{b'*}\rangle : \mathbf{S}_\Pi(0). \quad (43)$$

For a vertically incident plane *SH*-wave with a unit displacement in a direction $\mathbf{e}_x$ at the surface of the halfspace, $\underline{\mathbf{S}}_\Pi(0)$ is equal to $[\tilde{\Pi}^T(0) \otimes \tilde{\Pi}^T(0)] : \mathbf{e}_x \otimes \mathbf{e}_x$. The solution $\mathbf{S}_\Pi(0)$ of Eq. (43) allows to obtain the correlation $\tilde{\mathbf{S}}(\mathbf{k})$ in the wavenumber domain from Eq. (42). Integrating both the left and right hand sides of this equation over all wavenumbers and multiplying by $1/4\pi^2$ give the correlation $\mathbf{S}(\mathbf{x}''-\mathbf{x}')$ at $\mathbf{x}''=\mathbf{x}'$:

$$\mathbf{S}(0) = \underline{\mathbf{S}}(0) + \rho_{sc}\left[\frac{1}{4\pi^2}\int_{\mathbb{R}^2} \tilde{\underline{\mathbf{U}}}^G(\mathbf{k})\tilde{\Pi}(-\mathbf{k}) \otimes \tilde{\underline{\mathbf{U}}}^{G*}(\mathbf{k})\tilde{\Pi}(\mathbf{k})d\mathbf{k}\right] :\langle \mathbf{K}^{b'} \otimes \mathbf{K}^{b'*}\rangle : \mathbf{S}_\Pi(0). \quad (44)$$

The solution $\mathbf{S}_\Pi(0)$ is also used to calculate the correlation $\langle \mathbf{u}_b(\mathbf{x},0) \otimes \mathbf{u}_b^*(\mathbf{x},0)\rangle$ of the displacements at the interface Γ between a scatterer and the halfspace:

$$\langle \mathbf{u}_b(\mathbf{x},0) \otimes \mathbf{u}_b^*(\mathbf{x},0)\rangle = \left[\frac{1}{S_b}\int_\Gamma \Psi(\mathbf{x}) \otimes \Psi(\mathbf{x})dS\right] :\langle (\mathbf{K}^b + \mathbf{K}^s)^{-1}\mathbf{K}^s \otimes [(\mathbf{K}^b + \mathbf{K}^s)^{-1}\mathbf{K}^s]^*\rangle : \mathbf{S}_\Pi(0). \quad (45)$$

In the following, the correlation length is assumed to be large compared to the footprint of the scatterers, so that $\mathbf{S}(\mathbf{x}''-\mathbf{x}')$ approaches a constant value for $\mathbf{x}''$ and $\mathbf{x}'$ on Γ. As a result, only a 3×3 submatrix of $\mathbf{S}_\Pi(0)$ remains and needs to be calculated from Eq. (43) and accounted for in Eqs. (44) and (45).

TABLE I. The dynamic soil characteristics. Thickness t, shear wave velocity C_s, longitudinal wave velocity C_p, density ρ, and material damping ratio β of the soil layers.

Layer	t (m)	C_s (m/s)	C_p (m/s)	ρ (kg/m^3)	β
1	5	91	1447	1200	0.02
2	8	30	1447	1100	0.02
3	14	55	1233	1100	0.02
4	8	80	1267	1200	0.02
5	8	202	1442	1400	0.02
6	8	131	1472	1400	0.02
7	5	404	1786	1500	0.01
8	10	253	1588	1500	0.01
9	38	434	1746	1700	0.01
10	20	666	1965	1700	0.01
11	8	434	1771	1700	0.01
12	10	929	1935	1900	0.01
13	22	505	1776	1800	0.01
14	14	677	2084	1800	0.01
15	∞	1132	2522	2000	0.01

V. SEISMIC RESPONSE OF A CITY

A. The dynamic characteristics of the layered halfspace

In the following, the multiple scattering between buildings is studied for the case of a plane, vertically incident *SH*-wave, with dynamic soil and building characteristics corresponding to the case of Mexico City. The soil is modeled as a horizontally layered halfspace. Table I summarizes the layer thickness t, the shear wave velocity C_s, the longitudinal wave velocity C_p, the density ρ, and the material damping ratio β of the soil model. Material damping is modeled in the frequency domain by complex Lamé coefficients $\lambda(1+2\beta i)$ and $\mu(1+2\beta i)$, where β is the hysteretic material damping ratio. In order to assess the effect of multiple reflections at the interfaces between layers in the halfspace, the site transfer function is calculated for a vertically incident *SH*-wave. The site transfer function is the ratio between the response at the free surface of the layered halfspace in Table I and the response at the free surface of a homogeneous halfspace with the dynamic soil characteristics of the underlying halfspace (layer 15 in Table I). Figure 2 shows the site transfer function as a function of the circular frequency ω of the incoming wave. The presence of the upper, softer layers results in a

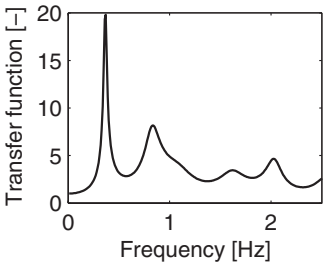


FIG. 2. Site transfer function.

first site resonance frequency with an amplification factor of 20 near 0.4 Hz. Higher site resonance frequencies are observed at 0.8, 1.6, and 2.0 Hz.

B. The dynamic characteristics of the scatterers

The buildings are modeled as elastodynamic cylinders uniformly distributed at the soil's surface. All buildings have a footprint $S_b=625\ m^2$, with a radius $R=14.1$ m. The areal density ρ_{sc} of the buildings is η/S_b, where η is the ratio of the built and total surface. A value of $\eta=0.06$ is considered, corresponding to a density $\rho_{sc}=1.02\times10^{-4}\ \mathrm{m}^{-2}$. The number of floors of the buildings follows the distribution in Fig. 3(a) for the Roma Norte district in Mexico City.[18] The height of the buildings equals $h=2.90n$. The mass m_b of the buildings is estimated as $0.36n\rho_bS_b$, with $\rho_b=2500\ \mathrm{kg}/m^3$ and 0.36 m an equivalent floor slab thickness. The mass is uniformly distributed over the cylinder.

The natural frequency of the first horizontal mode of a building fixed at its base is estimated from the number of floors n as $10/n$ using a common rule of thumb. In order to account for deviations between buildings with the same number n of floors, the natural frequencies are equally distributed over $\{10/n+1, 10/n+0.5, 10/n, 10/n-0.5, 10/n-1\}$. Since the city quarter consists of buildings with 6–16 floors, the fundamental resonance frequencies are between 0.6 and 2.0 Hz [Fig. 3(b)]. Resonance frequencies ω_m of higher modes are obtained applying a factor $2m-1$ on the fundamental resonance frequency. The corresponding mode shapes $\varphi_m(\mathbf{x})$ are the horizontal modes of a shear building. A uniform shear-beam continuum provides a good approximation of the horizontal response when the building has many identical stories.[37] The equation that governs shear wave propagation through the structure is similar to the one for axial wave propagation in a bar. The mode shapes $\varphi_m(\mathbf{x})$ for the horizontal motion in the direction $\mathbf{e}_x$ are therefore[37]

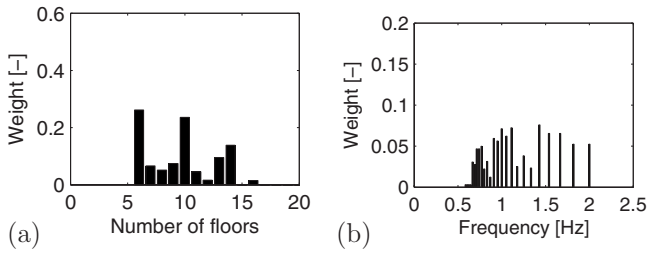


FIG. 3. Distribution of (a) the number of floors of the buildings and (b) the fundamental resonance frequencies.

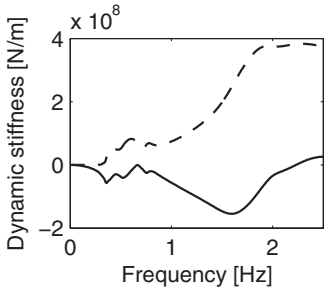


FIG. 4. Real (solid line) and imaginary parts (dashed line) of the first diagonal element $\langle[\mathbf{K}^{b'}]_{11}\rangle$ of the mean modified dynamic stiffness matrix.

$$\varphi_m(z)=\left\{\sqrt{\frac{2}{m_b}}\sin\left((2m-1)\frac{\pi z}{2h}\right),0,0\right\}^T. \tag{46}$$

The horizontal motion in the direction $\mathbf{e}_y$ is represented by similar mode shapes. The damping ratio ζ_m has a value of 0.05 for all buildings. These dynamic characteristics are used to calculate the dynamic stiffness $\mathbf{K}^b$ of the building using Eqs. (9)–(11) and (18). The dynamic stiffness $\mathbf{K}^s$ of the halfspace for a single scatterer is computed according to Eq. (13) using a boundary element method[26,27] based on the Green's functions of a layered elastic halfspace. The Green's functions are calculated by a reflection-transmission method.[33] The modified dynamic stiffness $\mathbf{K}^{b'}$ is computed from the dynamic stiffness $\mathbf{K}^b$ of the building and the dynamic stiffness $\mathbf{K}^s$ of the halfspace according to Eq. (26).

For a vertically incident *SH*-wave with a polarization according to $\mathbf{e}_x$, the mean response $\underline{u}_x(0)$ only depends on the first diagonal element $\langle[\mathbf{K}^{b'}]_{11}\rangle$ of the modified dynamic stiffness. Figure 4 shows the element $\langle[\mathbf{K}^{b'}]_{11}\rangle$ as a function of the circular frequency ω. At low frequencies, the dynamic stiffness is primarily real and scales as $-\omega^2$. In this frequency range, the buildings act as an additional mass at the soil's surface. At higher frequencies, the imaginary part becomes larger with local maxima at the natural frequencies of the buildings, coupled to the soil. The natural frequencies are slightly lower than those of a building fixed at its base [Fig. 3(b)], due to the additional flexibility provided by the soil.

C. Modification of the site response due to a plane *SH*-wave

The amplitude $u_{\mathrm{ffx}}(z)$ of the *SH*-wave at the free surface Γ_s of the homogeneous layered halfspace Ω_s is obtained as the product of the site transfer function (Fig. 2) and the frequency domain representation of a Ricker pulse centered at 0.75 Hz. The Ricker pulse therefore corresponds to the response at the free surface of a homogeneous halfspace with the properties of the underlying halfspace for the same incoming wave. Figure 5(a) compares the amplitude $u_{\mathrm{ffx}}(0)$ of the free field motion and the mean site response $\underline{u}_x(0)$ at the soil's surface. The maxima near 0.4 and 0.8 Hz of the response of the homogeneous layered halfspace correspond to the first and second site resonance frequencies. The higher resonance frequencies in the site transfer function (Fig. 2) are suppressed by the multiplication with the Ricker pulse and not observed in the response of the homogeneous layered

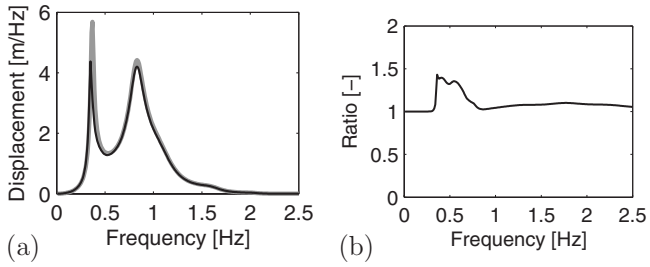


FIG. 5. (a) Comparison of the modulus of the mean (black line) response and the modulus of the response of the homogeneous layered halfspace (gray line). (b) Ratio of the mean square modulus of the response and the square modulus of the mean site response in the direction $\mathbf{e}_x$.

halfspace [Fig. 5(a)]. Comparing the mean response to the response of the homogeneous halfspace shows that the first site resonance frequency has shifted to a lower value due to the additional mass of the buildings. At higher frequencies, the mean site response is slightly smaller.

Figure 5(b) shows the ratio of the mean square modulus and the square modulus of the response in the direction $\mathbf{e}_x$. The mean square modulus is obtained from Eq. (43) for the filtered correlation $\mathbf{S}_{\Pi}(0)$ and a subsequent evaluation of the correlation $\mathbf{S}(0)$ according to Eq. (44). At low frequencies, the ratio is close to unity and the total site response is equal to the mean site response. In this frequency range, the wavelength in the soil is large compared to the building dimensions and the presence of the buildings results in a uniform modification of the site response. At higher frequencies, the ratio is larger due to the contribution of waves scattered by the buildings. The scattered waves are not accounted for in the mean response due to phase incoherence across the ensemble. The reduction in the mean site response (Fig. 5(a)) therefore corresponds to an apparent loss of energy. The observed loss of coherency shows that the buildings induce a spatial variation in the site response in this frequency range.

D. Influence of the scatterers

Figures 6(a)–6(c) compare the root mean square modulus of the displacements in the directions $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ to the modulus of the wavefield $\mathbf{u}_{\mathrm{ff}}(\mathbf{x},0)$ at the free surface Γ_s of the homogeneous layered halfspace Ω_s. The response is similar in the direction $\mathbf{e}_x$ [Fig. 6(a)], apart from the small shift of the first site resonance frequency that was also observed for the mean response [Fig. 5(a)]. In the directions $\mathbf{e}_y$ and $\mathbf{e}_z$, however, the wavefield at the free surface Γ_s of the homogeneous layered halfspace Ω_s is zero, whereas the waves scattered by the buildings contribute to these components of the wavefield. These transverse components are slightly smaller than the incident wavefield but have a similar order of magnitude.

Figure 7 compares the root mean square value of the time history of the displacements in the directions $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ which are obtained by a double inverse Fourier transform of the correlation in the frequency domain. Without buildings, the response in the direction $\mathbf{e}_x$ [Fig. 7(a)] shows an undulated behavior, corresponding to an exponentially decaying harmonic signal at the site resonance frequency. In the presence of buildings, a slightly smaller response is observed at early times and a higher response at later times. At early times, energy is transferred to the buildings, which is returned to the soil at later times. The undulated behavior is no longer observed at later times as the response mainly consists of scattered waves, with phase incoherence across the ensemble. These scattered waves are the only source of the response in the directions $\mathbf{e}_y$ and $\mathbf{e}_z$. Compared to the response in $\mathbf{e}_x$ [Fig. 7(a)], a slight time delay is observed corresponding to the time required for a building to retransmit energy into the soil after excitation by a seismic wave. The time scale is given by the dwell time t_{dwell}, estimated as $1/\omega_{\mathrm{ssi}}\zeta_{\mathrm{ssi}}$ (Ref. 25) from the natural frequency ω_{ssi} of the building coupled to the soil and the damping ratio ζ_{ssi} corresponding to radiation damping.

E. Influence of the scatterer footprint

In the following, the influence of the footprint S_b is investigated for a constant ratio $\eta=0.06$ of the built and total surface. Figure 8 shows the root mean square value of the frequency content for buildings with a footprint $S_b=312.5\ \mathrm{m}^2$ and $2500\ \mathrm{m}^2$, respectively. Apart from a smaller peak at the site resonance frequency for the smaller footprint, the response in $\mathbf{e}_x$ (Fig. 8) is relatively little affected by the difference in S_b. The shift of the first site resonance frequency is the same in both cases, as the ratio of the built and total surface and, therefore, the additional mass at the soil's surface is the same. The response in the directions $\mathbf{e}_y$ and $\mathbf{e}_z$ is much more affected by the change in the footprint S_b. The smaller footprint generally results in a smaller contribution of scattered waves to the response.

F. Influence of the scatterer internal damping

Figure 9 shows the root mean square value of the frequency content for buildings with a damping ratio $\xi=0.01$ and $\xi=0.20$, respectively. The value of 0.20 is unrealistically high for a building, but has been chosen to illuminate the role of energy dissipation in the buildings. When energy is

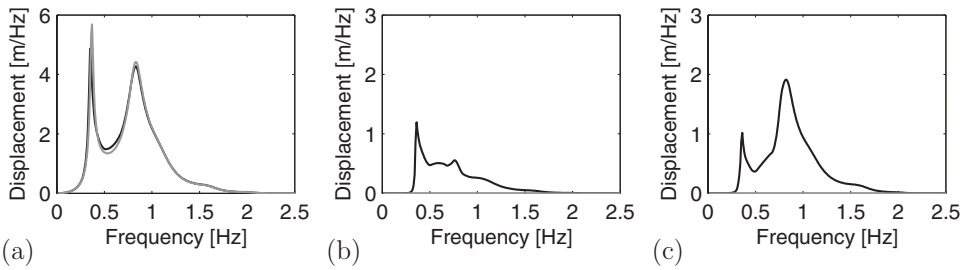


FIG. 6. Comparison of the root mean square modulus of the frequency contents (a) $\sqrt{\langle |u_x(\omega)|^2 \rangle}$, (b) $\sqrt{\langle |u_y(\omega)|^2 \rangle}$, and (c) $\sqrt{\langle |u_z(\omega)|^2 \rangle}$ for an areal density $\rho_{sc}=1.02\times10^{-4}\ \mathrm{m}^{-2}$ (black line) and the modulus of the wavefield at the free surface of the homogeneous layered halfspace (gray line).

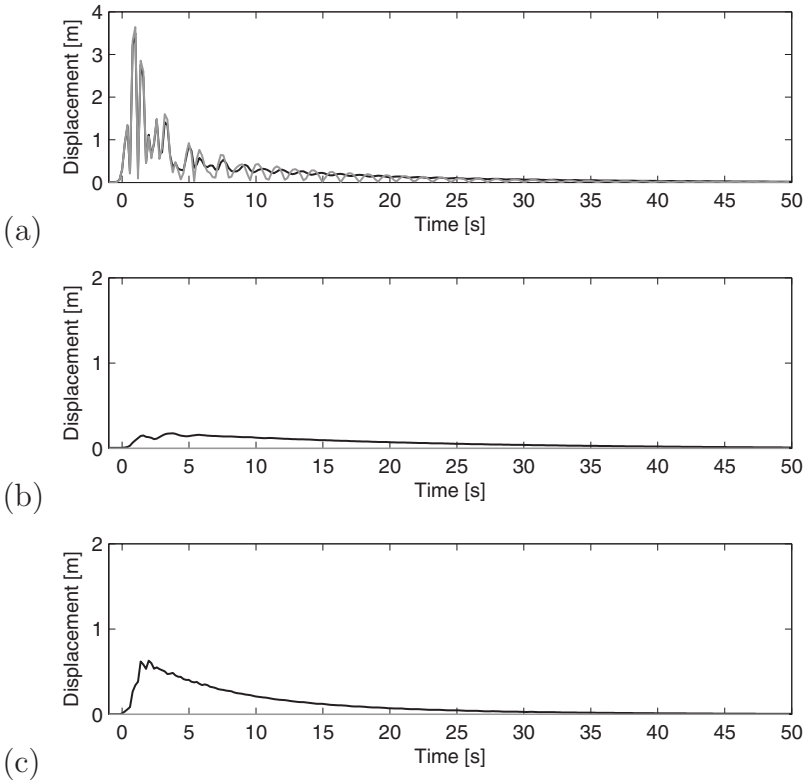


FIG. 7. Comparison of the root mean square time histories (a) $\sqrt{\langle|u_x(t)|^2\rangle}$, (b) $\sqrt{\langle|u_y(t)|^2\rangle}$, and (c) $\sqrt{\langle|u_z(t)|^2\rangle}$ for an areal density $\rho_{sc}=1.02\times10^{-4}$ m^{-2} (black line) and the absolute value of the wavefield at the free surface of the homogeneous layered halfspace (gray line).

transferred to the buildings, less energy is absorbed for the lower damping value, while more energy is radiated back into the soil. The response is only slightly affected, however, by the change in the damping ratio. This is due to the fact that energy is mainly absorbed by radiation and material damping in the soil.

VI. COMPARISON TO RESULTS FOR A CITY SAMPLE

A. Coupled FE/BE solution for a collection of buildings

The configurationally averaged field quantities are now compared to the results of a coupled finite element/boundary element method, where the buildings are modeled with finite elements and coupled to the layered halfspace model for the soil in Table I using the boundary element method. In order to determine the seismic response of the coupled system, a similar strategy is followed as in the single scatterer case (Sec. II). The displacements of each building k are decomposed according to Eq. (4) into the rigid body modes $\Psi^k(\mathbf{x})$ and the structural modes $\Phi^k(\mathbf{x})$ of the building fixed at the interface Γ_k between building k and the halfspace. The corresponding generalized coordinates are $\mathbf{c}^k$ and $\mathbf{q}^k$. The displacement field $\mathbf{u}_s(\mathbf{x})$ in the layered halfspace is decomposed into the wavefield $\mathbf{u}_{\mathrm{ff}}(\mathbf{x})$ in the homogeneous layered halfspace Ω_s without scatterers, the locally diffracted wavefield

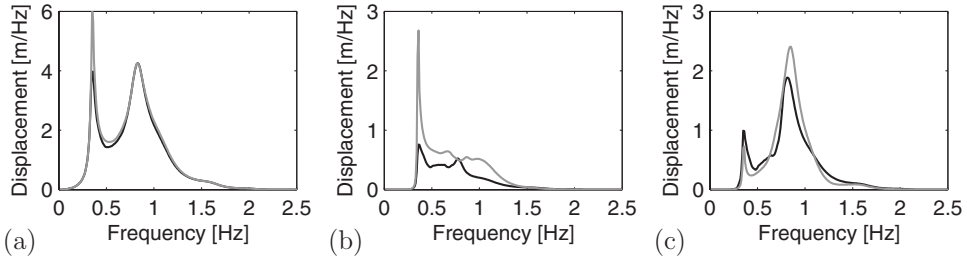


FIG. 8. Comparison of the root mean square modulus of the frequency contents (a) $\sqrt{\langle|u_x(\omega)|^2\rangle}$, (b) $\sqrt{\langle|u_y(\omega)|^2\rangle}$, and (c) $\sqrt{\langle|u_z(\omega)|^2\rangle}$ for the case of a city composed of buildings with a footprint $S_b=312.5$ m^2 (black line) and $S_b=2500$ m^2 (gray line).

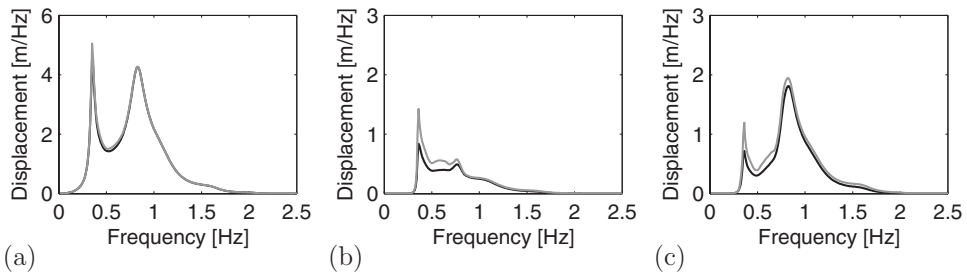


FIG. 9. Comparison of the root mean square modulus of the frequency contents (a) $\sqrt{\langle|u_x(\omega)|^2\rangle}$, (b) $\sqrt{\langle|u_y(\omega)|^2\rangle}$, and (c) $\sqrt{\langle|u_z(\omega)|^2\rangle}$ for the case of a city composed of buildings with an internal damping $\xi=0.01$ (black line) and $\xi=0.20$ (gray line).

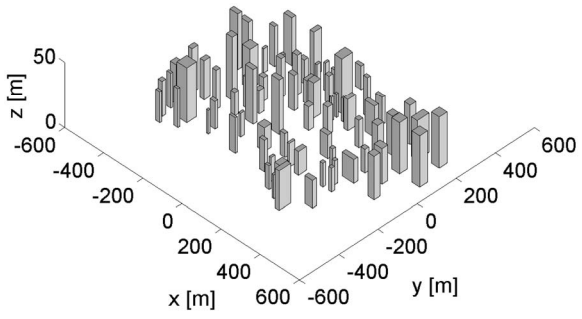


FIG. 10. The city configuration.

$\mathbf{u}_{d0}(\mathbf{x})$, that cancels the wavefield $\mathbf{u}_{\mathrm{ff}}(\mathbf{x})$ at every interface Γ_k, and the wavefields $\mathbf{u}_{dm}^k(\mathbf{x})$ scattered by the rigid body modes $\boldsymbol{\psi}_m^k(\mathbf{x})$ of the scatterers:[19,20]

$$\mathbf{u}_s(\mathbf{x}) = \mathbf{u}_{\mathrm{ff}}(\mathbf{x}) + \mathbf{u}_{d0}(\mathbf{x}) + \sum_{k=1}^{N_b} \sum_{m=1}^{6} \mathbf{u}_{dm}^k(\mathbf{x}) c_m^k, \tag{47}$$

where N_b is the number of buildings considered. The scattered wavefields $\mathbf{u}_{dm}^k(\mathbf{x})$ are by definition equal to $\boldsymbol{\psi}_m^k(\mathbf{x})$ for $\mathbf{x} \in \Gamma_k$ and zero for $\mathbf{x} \in \Gamma_l (l \neq k)$. A weak variational formulation of the equilibrium at $\cup \Gamma_k$ leads to

$$[\mathbf{K}^s + \mathbf{K}^b]\mathbf{c} = \mathbf{F}^s, \tag{48}$$

where $\mathbf{c}$ collects the generalized coordinates $\mathbf{c}^k$, and $\mathbf{K}^b$ is a block diagonal matrix with the dynamic stiffness matrices $\mathbf{K}^{bk}$ in Eq. (18) on its diagonal. The dynamic stiffness matrix $\mathbf{K}^s$ of the soil is a fully populated matrix composed of the following block matrices:

$$[\mathbf{K}^s]_{lk} = \int_{\Gamma_l} [\mathbf{T}_d^k(\mathbf{x})]^T \boldsymbol{\Psi}^l(\mathbf{x}) dS, \tag{49}$$

where $\mathbf{T}_d^k(\mathbf{x})$ collects the tractions $\mathbf{t}_s(\mathbf{u}_{dm}^k)(\mathbf{x})$. The vector $\mathbf{F}^s$ on the right hand side of Eq. (48) contains the seismic loading $\mathbf{F}^{sk}$ in Eq. (15) for each building. Solving Eq. (48) gives the generalized coordinates $\mathbf{c}^k$ of the modes $\boldsymbol{\psi}_m^k(\mathbf{x})$. The displacement field $\mathbf{u}_s(\mathbf{x})$ in the soil is obtained subsequently from Eq. (47).

B. Results

This methodology has been used to calculate the response of 77 buildings positioned on an area of 783 $\times$1006 m^2, as shown in Fig. 10. The ratio η of the built and total surface and the statistics of the number of building floors are the same as in Sec. V. In this model, however, the footprint of the buildings varies between 100 m^2 for four buildings with 6 floors and 2500 m^2 for a building with 14 floors. The previously considered value S_b=625 m^2 is the mean footprint. The dynamic stiffness matrix $\mathbf{K}^{bk}$ of the buildings is computed by means of the finite element method (FEM), using the MATLAB Structural Dynamics Toolbox.[38] The columns and beams of the buildings are modeled by beam elements, while plate elements are used for the floors of the buildings. All structural modes $\varphi_m(\mathbf{x})$ with a natural frequency below 30 Hz are accounted for. Both the dynamic stiffness matrix $\mathbf{K}_s$ of the soil in Eq. (49) and the seismic loading $\mathbf{F}^s$ in Eq. (15) are computed by a boundary element method[26] (BEM) with the computer program MISS.[27] The boundary element program uses the Green's functions of a horizontally layered halfspace, so that the radiation of waves into the layered halfspace and multiple wave scattering between the buildings are properly accounted for. The maximum element size in the discretization of the interfaces Γ_k is chosen as 5 m based on the results of a convergence study.[20]

Figure 11 compares the root mean square modulus of the frequency content as obtained by the Foldy–Ladder approximation and the result of the coupled FEM/BEM. Whereas all previous results concern the exciting field as obtained by the Foldy approach,[10,11] the response at the interface between the buildings and the soil is now considered. These results are calculated from the solution of Eq. (43) and a subsequent evaluation of Eq. (45). Assuming ergodicity, the spatially averaged field quantities for this city configuration can be compared to the ensemble averages from the present analysis. As the size of the city is of the same order of magnitude as the wavelength in the soil at low frequencies, it can be anticipated that its finite size will affect the results. The results of a second calculation by means of the Foldy–Ladder approximation with a smaller areal density $\rho_{sc}/16$ are therefore shown as well. The shaded area on the figures corresponds to the results for all buildings of the city quarter and illustrates the scatter of the response. Generally speaking, the response in the direction $\mathbf{e}_x$ is relatively well predicted, while in the directions $\mathbf{e}_y$ and $\mathbf{e}_z$, it is overestimated, particularly near the site resonance frequency and in the low frequency range. The overestimation of the contribution of scattered waves may be due to the finite size of the city that affects all buildings at low frequencies, and buildings at the city limits

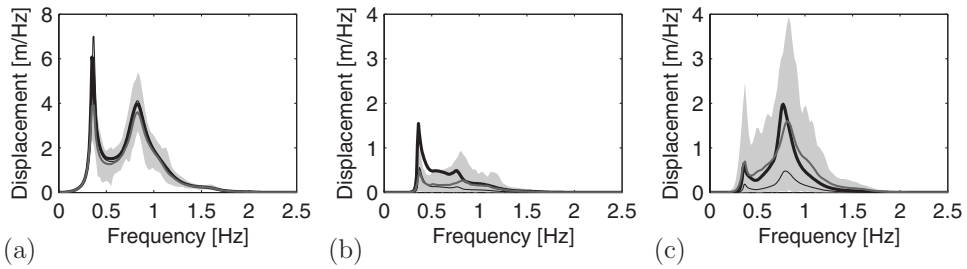


FIG. 11. Comparison of the root mean square modulus of the frequency contents (a) $\sqrt{\langle |u_x(\omega)|^2 \rangle}$, (b) $\sqrt{\langle |u_y(\omega)|^2 \rangle}$, and (c) $\sqrt{\langle |u_z(\omega)|^2 \rangle}$ of the response at the building-soil interface as obtained with the Foldy–Ladder approximation for an areal density $\rho_{sc}=1.02\times10^{-4}$ m^{-2} (thick black line) and $\rho_{sc}=6.39\times10^{-6}$ m^{-2} (thin black line) with the spatially averaged values for a particular city configuration (dark gray line). The envelope of all results for the particular city configuration is superimposed (light gray area).

at higher frequencies. Furthermore, the results in Fig. 8 show that the site response and, therefore, the response of the buildings are sensitive to the footprint of the buildings. Taking into account the difference between the building models, foundation areas, and the finite size of the city in the coupled FEM/BEM calculation, the agreement is reasonable.

VII. CONCLUSION

Within the frame of the present paper, the Foldy equation for the mean field and the Foldy–Ladder approximation of the Bethe–Salpeter equation are used to study the multiple scattering by randomly distributed finite-sized resonant scatterers at the surface of a horizontally layered elastic halfspace. Based on the solution in the single scatterer case, the configurationally averaged response and correlation are calculated for a vertically incident plane *SH*-wave. This solution is used to analyze the seismic response of an urban site, where each building acts as a scatterer at the soil's surface. In the low frequency range, where the wavelength in the soil is large with respect to the dimensions of the buildings, the presence of the buildings results in a uniform modification of the site response. The site response is therefore accurately characterized by the mean field, as there is no loss of coherency in this frequency range. At higher frequencies, the mean square response becomes substantially larger than the mean response, due to the fact that waves scattered by the buildings contribute to the seismic response of the site. The solution is compared to results for a city quarter, where multiple scattering between the buildings is fully accounted for.

ACKNOWLEDGMENT

G.L. is Postdoctoral Fellow of the Fund for Scientific Research-Flanders.

[1]C. Soize, "A model and numerical method in the medium frequency range for vibroacoustic predictions using the theory of structural fuzzy," J. Acoust. Soc. Am. **94**, 849–865 (1993).
[2]R. Lyon, "Statistical energy and structural fuzzy," J. Acoust. Soc. Am. **97**, 2878–2881 (1995).
[3]A. Pierce, V. Sparrow, and D. Russel, "Fundamental structural-acoustic idealizations for structures with fuzzy internals," J. Vibr. Acoust. **117**, 339–348 (1995).
[4]M. Strasberg and D. Feit, "Vibration damping of large structures induced by attached small resonant structures," J. Acoust. Soc. Am. **99**, 335–344 (1996).
[5]R. Weaver, "The effect of an undamped finite degree of freedom fuzzy substructure: Numerical solutions and theoretical discussion," J. Acoust. Soc. Am. **100**, 3159–3164 (1996).
[6]R. Weaver, "Mean and mean-square responses of a prototypical master/fuzzy structure," J. Acoust. Soc. Am. **101**, 1441–1449 (1997).
[7]R. Weaver, "Equipartition and mean-square responses in large undamped structures," J. Acoust. Soc. Am. **110**, 894–903 (2001).
[8]R. Weaver, "Multiple-scattering theory for mean responses in a plate with sprung masses," J. Acoust. Soc. Am. **101**, 3466–3747 (1997).
[9]R. Weaver, "Mean-square responses in a plate with sprung masses, energy flow and diffusion," J. Acoust. Soc. Am. **103**, 414–427 (1998).
[10]L. Foldy, "The multiple scattering of waves. I. General theory of isotropic scattering by randomly distributed scatterers," Phys. Rev. **67**, 107–119 (1945).
[11]P. Waterman and R. Truell, "Multiple scattering of waves," J. Math. Phys. **2**, 512–537 (1961).
[12]U. Frisch, "La propagation des ondes en milieu aléatoire et les équations stochastiques. Première partie (Wave propagation in random media and stochastic equations. First part)," Ann. Astrophys. **29**, 645–682 (1966).
[13]U. Frisch, "La propagation des ondes en milieu aléatoire et les équations stochastiques. II. Applications (Wave propagation in random media and stochastic equations. II. Applications)," Ann. Astrophys. **30**, 565–601 (1967).
[14]A. Lagendijk and B. van Tiggelen, "Resonant multiple scattering of light," Phys. Rep. **270**, 143–215 (1996).
[15]F. Chavez-Garcia and P.-Y. Bard, "Site effects in Mexico City eight years after the September 1985 Michoacan earthquakes," Soil. Dyn. Earthquake Eng. **13**, 229–247 (1994).
[16]A. Wirgin and P.-Y. Bard, "Effects of buildings on the duration and amplitude of ground motion in Mexico City," Bull. Seismol. Soc. Am. **86**, 914–920 (1996).
[17]D. Clouteau and D. Aubry, "Modification of the ground motion in dense urban areas," J. Comput. Acoust. **9**, 1659–1675 (2001).
[18]P. Guéguen, P.-Y. Bard, and F. Chavez-Garcia, "Site-city seismic interaction in Mexico City-like environments: An analytical study," Bull. Seismol. Soc. Am. **92**, 794–811 (2002).
[19]N. Mezher, D. Clouteau, and O. Ishizawa, "Modélisation numérique de l'intéraction multiple sol-structures (Numerical modelling of multiple soil-structure interaction)," in *Actes du Sixième Colloque National en Calcul des Structures*, edited by M. Potier-Ferry, M. Bonnet, and A. Bignonnet (Ecole Polytechnique, Giens, France, 2003), pp. 157–164.
[20]N. Mezher, "Modélisation numérique et quantification de leffet sismique site-ville (Numerical modelling and quantification of the seismic city-site effect)," Ph.D. thesis, Ecole Centrale de Paris, Châtenay-Malabry Cedex (2004).
[21]C. Boutin and P. Roussillon, "Assessment of the urbanization effect on seismic response," Bull. Seismol. Soc. Am. **94**, 251–268 (2004).
[22]A. Fernández-Ares and J. Bielak, "Interaction between earthquake ground motion and multiple buildings in urban regions," in Proceedings of the Third UJNR Workshop on Soil-Structure Interaction, edited by M. Celebi, M. Todorovska, I. Okawa, and M. Iiba (2004).
[23]J. Bielak, K. Loukakis, Y. Hisada, and C. Yoshimura, "Domain reduction method for three-dimensional earthquake modeling in localized regions, Part I: Theory," Bull. Seismol. Soc. Am. **93**, 817–824 (2003).
[24]M. Kham, J.-F. Semblat, P.-Y. Bard, and P. Dangla, "Seismic city-site interaction: Main governing phenomena through simplified numerical models," Bull. Seismol. Soc. Am. **96**, 1934–1951 (2006).
[25]G. Lombaert and D. Clouteau "The resonant multiple wave scattering in the seismic response of a city," Waves Random Complex Media **16**, 205–230 (2006).
[26]D. Aubry and D. Clouteau, "A regularized boundary element method for stratified media," in *Proceedings of the First International Conference on Mathematical and Numerical Aspects of Wave Propagation Phenomena*, edited by G. Cohen, L. Halpern, and P. Joly (SIAM, Philadephia, 1991), pp. 660–668.
[27]D. Clouteau, MISS Revision 6.2, Manuel Utilisateur (MISS Revision 6.2, User's Manual), Laboratoire de Mécanique des Sols, Structures et Matériaux, Ecole Centrale de Paris, 1999.
[28]M. Bonnet, *Boundary Integral Equation Methods for Solids and Fluids* (Wiley, Chichester, 1995).
[29]R. Apsel and J. Luco, "On the Green's functions for a layered half-space. Part II," Bull. Seismol. Soc. Am. **73**, 931–951 (1983).
[30]J. Luco and R. Apsel, "On the Green's functions for a layered half-space. Part I," Bull. Seismol. Soc. Am. **4**, 909–929 (1983).
[31]E. Kausel, *Fundamental Solutions in Elastodynamics: A Compendium* (Cambridge University Press, New York, 2006).
[32]E. Kausel and J. Roësset, "Stiffness matrices for layered soils," Bull. Seismol. Soc. Am. **71**, 1743–1761 (1981).
[33]B. Kennett, *Seismic Wave Propagation in Stratified Media* (Cambridge University Press, Cambridge, 1983).
[34]D. Photiadis, "Acoustics of a fluid-loaded plate with attached oscillators. Part I. Feynman rules," J. Acoust. Soc. Am. **102**, 348–357 (1997).
[35]R. Piessens, in *The Transforms and Applications Handbook*, The Electrical Engineering Handbook Series, edited by A. Poularikas (CRC, Boca Raton, FL, 1996).
[36]S. Rytov, Y. Kravtsov, and V. Tatarskii, *Principles of Statistical Radiophysics: Wave Propagation Through Random Media* (Springer-Verlag, Berlin, 1989), Vol. **4**.
[37]R. Clough and J. Penzien, *Dynamics of Structures* (McGraw-Hill, New York, 1975).
[38]E. Balmes, J.-P. Bianchi, and J.-M. Leclère, Structural Dynamics Toolbox User's Guide, SDTools, 2007.

A wave-based finite element analysis for acoustic transmission in fluid-filled elastic waveguides

Andrew T. Peplow
Hoare Lea Acoustics, 140 Aztec West Business Park, Bristol BS32 4TX, United Kingdom

(Received 2 February 2008; revised 26 January 2009; accepted 3 February 2009)

This paper describes an original numerical prediction technique developed for the analysis of coupled vibro-acoustic problems in fluid waveguides. Specifically it is a wave-based method that adopts a spectral element approach. Unlike the conventional element-based methods, this technique uses wave functions that satisfy the governing equations to describe the dynamic variables exactly. One advantage is that fine domain discretizations, used by element-based methods near the fluid-structure interface typically, are no longer required. Hence the resulting model sizes are much smaller than element-based methods yielding a more time-efficient prediction technique that may allow handling of mid-frequency applications. Another advantage is that dispersion relations between propagating and excitation wavenumbers are easily obtained, and an example to show this, as well as a discussion on how the wave-based prediction technique can be used for two coupled vibro-acoustic problems: a cavity with a non-reflecting boundary and a silencer duct problem, is presented. Its beneficial characteristics compared to element-based methods are demonstrated through the validation study and transmission loss examples.
 [DOI: 10.1121/1.3087430]

PACS number(s): 43.40.Rj, 43.20.Mv, 43.20.Bi [LLT] Pages: 2053–2063

I. INTRODUCTION

In this work a development of an acoustic wave-based technique for waveguides is presented. This technique now provides solutions to sound transmission through coupled structural-acoustic systems. The method used is a development to the acoustic spectral finite-element method (SFEM) by Peplow and Finnveden.[1] Essentially the wave-based method makes use of the exact acoustic solutions in one direction and an expansion in the cross-section consistent with that used for finite elements. The exact solutions are plane-wave-like functions. For problems such as acoustic radiation in a half-space, the exact solutions will be Bessel functions and spherical Bessel functions for two and three dimensions, respectively. Applying this strategy to a half-space leads to the known scaled boundary FEM that has been summarized for unbounded Helmholtz problems by Song.[2] Nevertheless, for waveguide problems, the approximation scheme used here is more or less similar to the partition of unity method (PUM) proposed by Babuška and Melenk[3] and developed by Laghrouche and Bettess.[4] The idea proposed is to employ a combination of polynomial functions and analytic functions for approximation without imposing any mesh-dependent conditions on the analytic functions. The benefit of the SFEM implemented here though over the PUM is only for geometries such as waveguides. That is, the analytic functions are chosen automatically through solving a system of homogeneous second order differential equations. However, extending the SFEM to radiation problems, mentioned above, will lead to difficulties described by Song[2] but will not be encountered by PUM. To be consistent and save confusion with earlier work[1] the wave-based method will be referred in the text henceforth as the SFEM.

Many noise control engineering designs encounter vibro-acoustic transmission problems. These problems involve the interaction between the vibrations of a mechanical structure and the sound field in its surroundings. At present, the most commonly used prediction techniques for solving complex duct problems numerically are finite-element-based methods. In the traditional FEMs, the problem domain is discretized into a finite number of small elements. All other standard element-based methods including boundary element methods also express the dynamic response variables in terms of simple but approximating shape functions. Because of this approximation, a sufficient number of elements per wavelength are required to obtain reasonable accuracy. With increasing frequency, wavelengths shorten, so that the number of elements must increase accordingly to maintain the same level of accuracy. Therefore, computational resources restrict the practical use of these element-based methods to low-frequency applications. However, in the past decade or so, a spectral wave-based prediction technique has been developed and has proven to be a successful tool for analyzing arbitrary sized vibro-acoustic problems in the low- and mid-frequency ranges.[5] The technique is based on an indirect semi-analytic approach where the dynamic response variables are described by a set of wave functions, which exactly satisfy the governing differential equations. These functions take the form $\exp(i\lambda x)$ and are thus dependent on the wavenumber λ. If the imaginary part of the wavenumber is zero the wave functions are propagating waves and these carry energy. Otherwise when the imaginary part of λ is non-zero the functions are evanescent waves and these do not transmit energy. The strategy for obtaining all wave functions, including exponentially increasing and decaying waves, is described in the work by Peplow and Finnveden.[1]

Previously the SFEM has been applied to a number of acoustic and structural waveguide problems. In particular, the method developed by Finnveden[5,6] can be viewed as a merger of the dynamic stiffness method and the FEM. The principle of the method is based on a variational formulation for non-conservative motion in the frequency domain. Use of a variational formulation for the spectral method provides a natural basis for approximations and a straightforward method for combination with standard finite elements. Indeed, the motivation behind this approach is to permit direct coupling with standard finite-element software. The end product allows non-standard problems in computational wave propagation to be resolved.

In the super-SFEM pursued here, the region of interest is decomposed into rectangular elements. A variational formulation for the underlying equations of motion is then approximated using trial functions that are conforming across element boundaries. The pressure field in each element is expressed as a combination of polynomials in the transversal z-direction with wave influence functions in the longitudinal x-direction. From the variational formulation eigenfunctions of a derived differential operator are chosen as the wave influence trial functions in the description of the acoustic pressure and displacement fields for the local *spectral element*. Local dynamic stiffness matrices for all the elements are then combined to form the global stiffness matrix for the complete problem and hence given a particular source the pressure and displacement fields may be found at any point within the fluid and structural regions. Since the element does not rely on any wavelength-dependent discretization in the longitudinal direction the SFEM can be applied to fairly short as well as very long wave transmission problems. In this way, only partial discretization is needed, and since these are smaller and more computationally efficient models, this technique allows also handling of mid-frequency applications.

Previously the SFEM technique has been used to study vibrations in beam frameworks[5,7] and for fluid-filled pipes.[6] Here the important problem for calculating propagating waves, as these are the only waves which transmit energy, is developed in Sec. V, which is identical to that analyzed by Fahy.[8] Coupled dispersion characteristics for fluid-filled cylindrical pipes have also been presented by Fuller and Fahy.[9] The authors used a numerical solution methodology in solving the coupled dispersion equation by performing studies for different values of shell thickness to radius ratio and also different material combinations. The behavior of various branches of the dispersion curve under the influence of these parameters was also discussed in the article by Pavic.[10] In his study of vibrational energy flow in a cylindrical shell the author also obtained the dispersion curves of a fluid-filled cylindrical shell. Similarly Finnveden[5] studied the effect of a flange on wave propagation down a fluid-filled elastic cylinder. This also included finding dispersion curves for the fluid-filled pipe through a finite-element discretization in the radial direction. Here we use a similar methodology to devise finite-element basis functions to generation wave influence functions for fluid structural elements. We study a simple validation case and a reactive flexible duct assemblage example. Dissipative devices such as acoustic linings or silencers are often used to reduce the noise levels. These work well in the mid- and high-frequency ranges but are less effective for low-frequency acoustics. In many applications the silencer device comprises a two-dimensional reactive silencer in which a flexible structure is attached to the external wall of the expansion chamber parallel to the axis of the inlet/outlet ducts. Controlling radiated noise by a membrane or plate at the interface between a silencer and the duct has been studied extensively by Huang and co-workers[11–15] where the stiffness of the air-back cavity has been employed as a controlling parameter. If, however the height of the membrane above the level of the inlet/outlet ducts can be varied, the device may be tuned by selecting the membrane height that gives the widest stop-band for a specified frequency.[11] In this paper there is no air-backed cavity; the radiation impedance of the plate is due to the interaction of the air duct mode and the *in vacuo* mode of the plate.

The basic concepts of the computational technique are briefly discussed for a vibro-acoustic problem defined in a bounded cavity domain. The performance of the new technique in terms of computational load is compared with existing FEMs through this validation example.

The purpose of the present study is to evaluate a new computational method for predicting sound transmission through flexible ducts with non-uniform but regular waveguides. The paper is organized as follows: In Sec. II the variational principle for non-conservative acoustic motion is stated. In Secs. III and IV a description of the trial functions is given and associated dynamic stiffness matrices are presented. Acoustic responses and displacements for the structure for various geometries are shown in Sec. V and some conclusions are given in Sec. VI.

II. ANALYSIS

The acoustical pressure and structural flexural displacement will be denoted by $p^*(\underline{x},t)$ and $w^*(\underline{x},t)$, respectively. The linearized acoustical wave equation, without sources, for the steady-state pressure assuming harmonic motion, $e^{-i\omega t}$, on suppressing the asterisks, then becomes

$$\nabla^2 p(\underline{x}) + k^2 p(\underline{x}) = 0, \quad \underline{x} \in \Omega_P, \tag{1}$$

where $k=\omega/c_0$ is the acoustic wavenumber, c_0 is the sound speed, i is the imaginary unit $i=\sqrt{-1}$, ∇^2 is the two-dimensional Laplacian operator, and coordinate vector $\underline{x}=(x,z)$. The physical system may be excited at a frequency $\omega=2\pi f$, through acceleration $\nabla_p=\rho_0\dot{\mathbf{u}}$ where ρ_0 is density of fluid and $\mathbf{u}$ represents the particle velocity. A typical fluid-filled rectangular region, Ω_F in the x-z plane, and a fluid-filled cavity backed by a thin, flexible plate Ω_P, on the right hand side, are both shown in Fig. 1(b). The origin is indicated and the plate, denoted by Γ_P, lies on the x-axis.

The steady-state normal displacement $w(\underline{x}_s)$ of the plate is governed by the dynamic equation

$$(D_s\nabla^4 - \rho_s t\omega^2)w(\underline{x}_s) = p(\underline{x}_s), \quad \underline{x}_s \in \Gamma_P, \tag{2}$$

where D_s denotes flexural rigidity, ρ_s is the mass per unit volume, t is plate thickness, and $(x,0)=\underline{x}_s \in \Gamma_P$. To uniquely

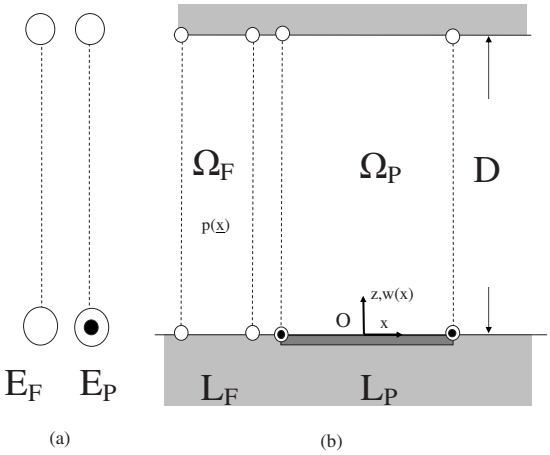


FIG. 1. Illustration of the two types of elements: (a) cross-sectional WFEs, acoustic (E_F) and coupled acoustic-structure elements (E_P), and elements defined over regions Ω_F and Ω_P of height D; (b) acoustic spectral finite element of width L_F (left) and coupled fluid-structure spectral finite element of width L_P (right).

define the displacement, four boundary conditions must be specified, that is, conditions are required at the two plate ends. These can model clamped, simply supported, or free plate ends. All combinations of the boundary conditions may be included in this analysis; however, for the model examples here, only clamped and simply supported ends are examined. Hence for the plate defined over length L_P: $\underline{x}_s \rightarrow x \in [-L_P/2, +L_P/2]$, two sets of boundary conditions at end-points $x_1 = -L_P/2$ and $x_2 = L_P/2$ are required, either simply supported (SS) or clamped (CC) at both ends, respectively,

$$w(x_1) = w(x_2) = 0, \quad \frac{d^2w}{dx^2}(x_1) = \frac{d^2w}{dx^2}(x_2) = 0 \tag{3}$$

or

$$w(x_1) = w(x_2) = 0, \quad \frac{dw}{dx}(x_1) = \frac{dw}{dx}(x_2) = 0. \tag{4}$$

If Γ_F denotes the boundary for the fluid-filled region, the acoustic boundary condition for absorbing locally reacting walls is given by

$$\frac{i}{\rho_0 \omega}\frac{\partial p}{\partial n}(\underline{x}) = \frac{p(\underline{x})}{\zeta}, \quad \underline{x} \in \Gamma_F, \tag{5}$$

where n denotes the normal outgoing direction and ζ is the surface impedance. To define an interface condition between the fluid and the structure, continuity of fluid particle velocity and structure velocity is specified as

$$\frac{i}{\rho \omega}\frac{\partial p}{\partial n}(\underline{x}_s) = -i\omega w(\underline{x}_s), \quad \underline{x}_s \in \Gamma_P. \tag{6}$$

Figure 1(b) shows a typical geometry of a plate enclosed by fluid-filled cavity used in the analysis.

The devised FEM here derives solutions to coupled problems defined through Eqs. (1)–(6). Briefly, the implementation of the numerical method is split into two stages: development of *waveguide finite elements* (WFEs) defined over cross-sections [Fig. 1(a)], and then construction of *super-spectral finite elements* defined over combinations of arbitrary length regions [Fig. 1(b)], which provide the dynamic stiffness matrices.

A. Variational statement

In what follows the displacement-pressure variables are converted to a displacement-velocity potential (w, ϕ) formulation for purposes of symmetry and likely non-dimensional convenience, $p(\underline{x}) = -i\rho_0 \omega \phi(\underline{x})$. A typical waveguide airway section is shown in Fig. 1(b). We assume that the waveguide is formed by a uniform constant-depth rectangular region of acoustic fluid surrounded by a sound absorbing lining or backed by a vibrating plate. The finite element formulation allows geometrical rectangular discontinuities such as stairsteps or silencers; however, for the WFE construction in this section, we assume that the waveguide is divided equally in the vertical direction throughout.

Equations within Ω_P and the continuity conditions at all interfaces may be expressed in a single weak variational statement. That is, the full problem may be reformulated in a convenient form as a stationary solution of the functional[6]

$$\begin{aligned} L_{\Omega_P} = \rho_0 \int_{\Omega_P} -k^2 \phi^a \phi + \nabla \phi^a \cdot \nabla \phi d\Omega_P - \rho_0 \int_{\Gamma_F} \phi^a \frac{\partial \phi}{\partial n} \\ + \phi \frac{\partial \phi^a}{\partial n} d\Gamma_F - \rho_0 \omega \int_{\Gamma_P} \phi^a w + w^a \phi d\Gamma_P \\ + \int_{\Gamma_P} \left(D_S \frac{d^2 w^a}{dx^2}\frac{d^2 w}{dx^2} - m_s \omega^2 w^a w \right) d\Gamma_P, \end{aligned} \tag{7}$$

where $\Omega_P = \Omega_F \cup \Gamma_P \cup \Gamma_F$. The superscripted symbol ϕ^a denotes the complex conjugate of the sound pressure in the adjoint system, similar to the original system with negative damping.[5] For ease of exposition the losses have been omitted in the finite element formulations in Secs. III and IV but are included in the numerical examples in Sec. V through complex valued acoustic wavenumber and flexural rigidity. For brevity fluid boundaries, Γ_F are assumed rigid so that integrals over these boundaries are also omitted. Modifications of the finite element matrices to include losses and boundary impedance are straightforward.[1] As shown in Ref. 5 and discussed in Refs. 1 and 6, the losses may be accounted by employing a variational principle similar to Hamilton's principle whereby, in the functionals describing kinetic and potential energies, the quadratic forms are replaced with symmetric bilinear forms and complex conjugates in the adjoint negatively damped system. With this approach the mechanisms of variational methods are retained for non-conservative systems. The ansatz above, Eq. (7), may now be used as a variational form for the acoustical and structural response solutions. The procedure for a finite-element idealization for the waveguide, and generation of suitable basis functions, is detailed in Sec. III.

III. FINITE ELEMENT MODELING

In fluid-filled flexible waveguides with constant geometrical cross-section, the solutions of the governing equations of motion are a combination of exponential terms and

cross-sectional modes that describe the propagation and transmission of energy. Thus, the axial x-dependence is separable from the cross-sectional z-dependencies. For the displacement and fluid motion a combination of polynomials is assumed. In the finite-element discretization a cross-section is divided so that material parameters are constant within each interval. Within one WFE, the acoustic and displacement fields may be represented in the coordinate system. A dimensionless $(1\times m)$ real-valued matrix of polynomials $[\hat{\mathbf{N}}(z)]=[N_1(z),N_2(z),\ldots,N_m(z)]$ is defined with the given coordinate functions as elements. A column vector $\mathbf{S}(x)$ of unknowns is defined as

$$S(x,z)=(w(x),\phi(x,z))^T=[\mathbf{N}(z)]\mathbf{S}(x), \tag{8}$$

where the matrix

$$[\mathbf{N}(s)]=\begin{bmatrix}1 & \mathbf{0}\\ 0 & \tilde{\mathbf{N}}(z)\end{bmatrix}, \tag{9}$$

$[\mathbf{N}(z)]$ is a $2\times(m+1)$ matrix, and $\mathbf{0}$ denotes a zero $(1\times m)$ row vector. Section III A describes a development of the variational approach for determining the WFEs $\mathbf{S}(x)$ by construction of a polynomial eigenvalue problem.

A. The coupled vibro-acoustic model

First the variational statement, Eq. (7), in Sec. II A is developed further to obtain *wave influence* basis functions $\mathbf{S}(x)$.

Consider the functional L_{Ω_P}, Eq. (7), for a single arbitrary region $\Omega_P=c\times d$. Substitution of expression (9) into the resulting form, where subscript $_x$ denotes x-derivative, yields the approximation

$$\begin{aligned}L_{\Omega_P}=&-k^2\int_d \mathbf{S}^a(x)[\mathbf{M}]\mathbf{S}(x)dx+\int_d \mathbf{S}^a(x)[\mathbf{K}_0]\mathbf{S}(x)dx\\&+\int_d \mathbf{S}^a(x)[\mathbf{K}_1]\mathbf{S}(x)dx+\int_d \mathbf{S}_x^a(x)[\mathbf{K}_2]\mathbf{S}_x(x)dx\\&+\int_d \mathbf{S}_{xx}^a(x)[\mathbf{K}_4]\mathbf{S}_{xx}(x)dx.\end{aligned} \tag{10}$$

Square $(m+1)\times(m+1)$ matrices are given by

$$[\mathbf{M}]=\begin{pmatrix}m_s c_0^2 & \mathbf{0}\\ \mathbf{0}^T & [\hat{\mathbf{M}}]\end{pmatrix},\quad [\mathbf{K}_0]=\begin{pmatrix}0 & \mathbf{0}\\ \mathbf{0}^T & [\hat{\mathbf{K}}_0]\end{pmatrix}, \tag{11}$$

$$[\mathbf{K}_1]=-\rho_0 c_0\omega\begin{pmatrix}0 & \mathbf{1}\\ \mathbf{1}^T & [\hat{\mathbf{0}}]\end{pmatrix},\quad [\mathbf{K}_2]=\begin{pmatrix}0 & \mathbf{0}\\ \mathbf{0}^T & [\hat{\mathbf{K}}_2]\end{pmatrix}, \tag{12}$$

$$[\mathbf{K}_4]=\begin{pmatrix}D_S & \mathbf{0}\\ \mathbf{0}^T & [\hat{\mathbf{0}}]\end{pmatrix}. \tag{13}$$

Note that all the matrices are real, square, and symmetric $(m+1)\times(m+1)$ matrices. In particular, $[\mathbf{M}]$ is regular positive definite and $[\mathbf{K}_4]$ is a real, rank-deficient symmetric $[\mathrm{rank}([\mathbf{K}_4])=1]$ matrix. Matrices $[\mathbf{K}_2]$, $[\mathbf{K}_1]$, $[\mathbf{K}_0]$, and $[\mathbf{M}]$ are given by

$$[\hat{\mathbf{K}}_2]=\rho_0\int_c [\mathbf{N}]^T[\mathbf{N}]\mathrm{d}z, \tag{14}$$

$$[\hat{\mathbf{K}}_0]=\rho_0\int_c [\nabla_z\mathbf{N}]^T[\nabla_z\mathbf{N}]\mathrm{d}z, \tag{15}$$

$$[\hat{\mathbf{M}}]=\rho_0\int_c [\mathbf{N}]^T[\mathbf{N}]\mathrm{d}z. \tag{16}$$

Note that $[\hat{\mathbf{K}}_1]$ is the coupling matrix and matrices $[\hat{\mathbf{M}}]$ and $[\hat{\mathbf{K}}_2]$ are identical.

B. Waveguide finite element modeling

A resulting set of ordinary differential equations to the Euler–Lagrange equations, which correspond to Eq. (10), may be found by taking an appropriate first variation

$$\left([\mathbf{K}_4]\frac{d^4}{dx^4}-[\mathbf{K}_4]\frac{d^2}{dx^2}+k[\mathbf{K}_1]+[\mathbf{K}_0]-k^2[\mathbf{M}]\right)\mathbf{S}(x)=0. \tag{17}$$

The system of equations, Eq. (17), has constant coefficients in the form of the real, symmetric $(m+1)\times(m+1)$ matrices. Hence, the solutions of the linear homogeneous system may be written as

$$\mathbf{S}(x)=\mathbf{\Phi}e^{i\lambda x}, \tag{18}$$

where $\mathbf{\Phi}$ is the column vector representing the nodal structural amplitudes, Φ_1, and $\Phi_2,\ldots,\Phi_{m+1}$ representing the velocity potential amplitudes. Under this assumption Eq. (17) reduces to a fourth order polynomial eigenvalue problem

$$(\lambda^4[\mathbf{K}_4]+\lambda^2[\mathbf{K}_2]+k[\mathbf{K}_1]+[\mathbf{K}_0]-k^2[\mathbf{M}])\mathbf{\Phi}=0. \tag{19}$$

Given a frequency ω or acoustic wavenumber k and collecting terms this expression may be abbreviated as

$$(\lambda^4[\mathbf{K}_4]+\lambda^2[\mathbf{K}_2]+[\bar{\mathbf{K}}_0])\mathbf{\Phi}=0, \tag{20}$$

where $[\bar{\mathbf{K}}_0]=-k^2[\mathbf{M}]+k[\mathbf{K}_1]+[\mathbf{K}_0]$. We seek all the eigenvalues or wavenumbers, λ; the leading matrix in the polynomial expansion, Eq. (20), is a "rank 1" singular matrix and hence the procedure for reducing the polynomial eigenvalue problem to a simple eigenvalue problem is not straightforward.[16] We shall see that a solution procedure is possible for Eq. (20) that yields eigenvalues in pairs, $\lambda^{\pm}=\pm\lambda$, indicating that pairs of eigenmodes result to Eq. (18) with the same phase speed propagating in the positive and negative axial directions. To resolve the quadratic eigenvalue problem (20) a dummy variable $\lambda^2\mathbf{\Phi}$ is introduced as discussed in Ref. 16. This extends the dimension of the eigenvalue problem to $2(m+1)\times 2(m+1)$ as

$$\left\{\begin{bmatrix}\bar{\mathbf{K}}_0 & 0\\ \mathbf{K}_2 & -\bar{\mathbf{K}}_0\end{bmatrix}-\lambda^2\begin{bmatrix}0 & \bar{\mathbf{K}}_0\\ -\mathbf{K}_4 & 0\end{bmatrix}\right\}\begin{Bmatrix}\lambda^2\mathbf{\Phi}\\ \mathbf{\Phi}\end{Bmatrix}=0. \tag{21}$$

The left hand matrix is invertible and the right hand matrix is rank-deficient equal to $(m+2)$. The generalized matrix eigenvalue problem above may be rewritten as standard

eigenvalue problem, in $1/\lambda^2$, by multiplying the full expression by the invertible matrix. It is clear then that the resulting eigenvalue problem has at most $(m+2)$ distinct, non-zero finite eigenvalues. That is, there are at least $(m+2)$ infinite eigenvalues, λ^2, to be discarded as non-physical. Since the standard eigenvalue problem is equivalent to the original matrix problem (21) the eigenspace is spanned by $(m+2)$ eigenvectors each of dimension $2(m+1)$. On inspection of Eq. (21) this matrix problem provides $2(m+2)$ positive and negative propagating wavenumbers $\lambda^{\pm}$ with corresponding eigenvectors $\Phi^{\pm}$ of length $(m+1)$. Note that the eigenvalue problem is not solved numerically as a simple eigenvalue problem described here. The resolution of the matrix eigenvalue problem, Eq. (21), itself may be achieved by a number of standard computational routines. In the present analysis a *QZ* algorithm was used as implemented in MATLAB 7.0.2. This allows a "rank deficient" generalized eigenvalue problem to be solved and produces a set of finite number of complex-valued eigenvalues λ_j^2, $j=1,\ldots,(m+2)$ with corresponding cross-sectional *mode shapes*.

IV. SUPER-SPECTRAL FINITE-ELEMENT MODELING

An approximate solution to the original problem is determined by selecting a solution $[w(x),\phi(x,z)]$ from a discrete set of basis functions determined by a finite-element discretization of a region Ω_P, Fig. 1(b). Construction of the set of basis functions, Eq. (8), and local dynamic stiffness matrix, defined over the region Ω_P as in Fig. 1(b), $x \in d$, where $-D \leqslant x \leqslant +D$ is described in Ref. 1 with minor amendments to take into account the structural displacement. By consideration of the solutions in Eq. (18) it is clear that each wave influence function may be written as a linear combination of the following functions:

$$S_{jk}(x) = \sum_{l=1}^{2(m+2)} \bar{\Phi}_{jl} E_l(x) A_{lk} s_k, \quad j = 1, \ldots, (m+1), \quad k = 1, \ldots, 2(m+2), \tag{22}$$

where entries $\bar{\Phi}_{jl}$ and $E_l(x)$ take the values of eigenvectors and exponential functions, respectively. Coefficients A_{lk} are determined by scaling the set of wave influence functions to unity as described by Peplow and Finnveden.[1] For the variational parameters, there corresponds m acoustic degrees of freedom leaving two degrees of freedom for the structure. For example, if the unknowns are denoted by the vectors $(W_{xx}, W, \phi_1, \ldots, \phi_m)$ then simply supported conditions (3) may be represented through the combination

$$S_{jk}(x) = \begin{cases} \displaystyle\sum_{l=1}^{2(m+2)} \bar{\Phi}_{jl} \Lambda_l^2 E_l(x) A_{lk} s_k \\ k = 1, \ldots, 2(m+2), \quad j = 1 \\ \displaystyle\sum_{l=1}^{2(m+2)} \bar{\Phi}_{jl} E_l(x) A_{lk} s_k \\ k = 1, \ldots, 2(m+2), \quad 2 \leqslant j \leqslant (m+1) \end{cases} \tag{23}$$

or through the unknowns $(W_x, W, \phi_1, \ldots, \phi_m)$ clamped conditions (4) may be represented using the combination

$$S_{jk}(x) = \begin{cases} \displaystyle\sum_{l=1}^{2(m+2)} \bar{\Phi}_{jl} \Lambda_l E_l(x) A_{lk} s_k \\ k = 1, \ldots, 2(m+2), \quad j = 1 \\ \displaystyle\sum_{l=1}^{2(m+2)} \bar{\Phi}_{jl} E_l(x) A_{lk} s_k \\ k = 1, \ldots, 2(m+2), \quad 2 \leqslant j \leqslant (m+1). \end{cases} \tag{24}$$

Hence, the Lagrangian and the local dynamic stiffness matrix, $[L]_{\log}$, for a certain super-spectral element may be written as

$$L_\Omega = \mathbf{s}^T [L]_\infty \mathbf{s}, \tag{25}$$

where

$$[L]_{\text{loc}} = [A]^T ([\bar{\Phi}]^T (-k^2[\mathbf{M}] + [\mathbf{K_0}] + k[\mathbf{K_1}])[\bar{\Phi}] + \langle \mathbf{\Lambda} \rangle [\bar{\Phi}]^T [\mathbf{K_2}] [\bar{\Phi}] \langle \mathbf{\Lambda} \rangle + \langle \mathbf{\Lambda^2} \rangle [\bar{\Phi}]^T [\mathbf{K_4}] [\bar{\Phi}] \langle \mathbf{\Lambda^2} \rangle) * [E_l]][A]. \tag{26}$$

where $*$ denotes the Schur product. The global dynamic stiffness matrix is assembled by calculated local dynamic stiffness matrices for each region, such as Ω_P in Fig. 1(b). For example, consider a uniform waveguide geometry consisting of three super-spectral elements, with $(m+2)$ degrees of freedom across each element cross-section. A $2(m+1) \times 2(m+1)$ generalized matrix eigenvalue problem is solved for the WFE and a $2(m+2) \times 2(m+2)$ local dynamic stiffness matrix (26) is generated for each super-spectral element. Enforcing continuity of velocity potential, displacement, and slope across neighbors and combining local dynamic stiffness matrices result in a $4(m+2) \times 4(m+2)$ frequency-dependent global dynamic stiffness matrix. Acoustic sources are modeled as either volume point sources or given mechanical sources such as normal velocities or accelerations defined on boundaries. These are introduced into the governing equations of motion via the variational formulation (7).

Evaluation of the dynamic stiffness matrix above applies to a simple super-spectral finite element for a given frequency, $\omega = 2\pi f$. The frequency-dependent matrix is banded, complex, and symmetric. For a combination of super-spectral finite elements, describing a general fluid domain, the corresponding dynamic stiffness matrix has a block diagonal structure; each block derived from the expression in Eq. (26).

The complete element formulation, for a single arbitrary length super-spectral element, costs little in terms of computational expenditure. Including assembling local element matrices, solving eigenvalue problems, and evaluating the global dynamic stiffness matrix, Eq. (26), this required approximately 0.04 s on a 1.3 GHz Pentium Centrino PC for a 13, $(m=11)$, degree-of-freedom super-spectral element. Accuracy of the method is similar to a one-dimensional standard FEM and is described in Ref. 1.

V. NUMERICAL EXAMPLES

In the following we illustrate the efficacy of the present approach through accurate solutions to simple canonical

propagation and transmission problems in water-filled waveguides. If not specified the wavespeed of the fluid was taken to be $c_0=1500$ ms^{-1} with density $\rho_0=1000$ kg m^{-3}. Zero damping is assumed for evaluation of dispersion relations but typical values used in the validation sections, Secs. V B and V C, are 0.001% and 1.0% for the fluid and structure, respectively. Dispersion relations for coupled systems with given propagating wave mode shapes are shown first followed by examples for forced responses where frequency values for excitation were not only chosen arbitrarily but to best illustrate each case. The form of the flexural rigidity of the steel plate D_s is given by EI where E is the modulus of elasticity $E=2\times10^{11}$ N/m^2 of and I is the moment of inertia $I=t^3/12(1-\nu^2)$ where t is the thickness of the plate, $\nu=0.3$ is Poisson's ratio, ρ_s is the mass per unit volume 7850 kg/m^3, and hence $m_s=\rho_s t$ is the mass per unit length of the beam. Note that the SFEM solutions found have been converted and presented as acoustic pressure values.

A. Dispersion relations

Dispersion relations between propagating wavenumbers and frequency provide important information about vibration and acoustical energy that may transmit through a coupled system. Useful by-products, then, of the SFEM are the calculation propagating wavenumbers, or the eigenvalues, λ (with zero imaginary part) also referred to as the propagating wavenumbers, and may be obtained from the matrix eigenvalue problem (19). For a given frequency the element matrices are assembled, Eqs. (11)–(13), yielding solutions of the quadratic eigenvalue problem through Eq. (19). Three cases are considered, a fluid-only case, a fluid coupled to a "heavy" steel structure, and a fluid coupled to a "light" steel structure. The parameters were chosen to highlight dominant features of the dispersion curves in the full coupled system. First the fluid-only problem, depicted by the left-hand plot in Fig. 1(a) and approximated using quartic polynomial shape functions, Eq. (9), defined over three elements in the z-direction resulted in 13×13 element matrices, Eqs. (11)–(13). For this case the uncoupled fluid-only eigenvalue problem, not shown here, was adopted; see Ref. 1 for details. Figure 2 shows the dispersion relations found from the numerical scheme against exact values shown by dashed lines. The right hand plots in Fig. 2 show the eigenvectors associated with the propagating modes at two frequencies, 2000 and 3500 Hz. The zero order (plane-wave) mode, depicted by the dashed line, and the first and second order modes, depicted by dashed-dotted and dotted lines, are clearly evident.

Now consider the situation including a structure depicted in right-hand plot of Fig. 1(a). Using the same element shape functions for the fluid domain the fully coupled eigenvalue problem (21) increases to a 28×28 generalized eigenvalue problem. Given two thickness of plates, $t=22.0$ mm and $t=4.0$ mm, the dispersion relations up to frequency 5.0 kHz are shown in Figs. 3 and 4.

In the coupled fluid-structure systems the appearance of dispersion curves, along the frequency axis, now shifts to the right in frequency as the plate thickness decreases. Notably

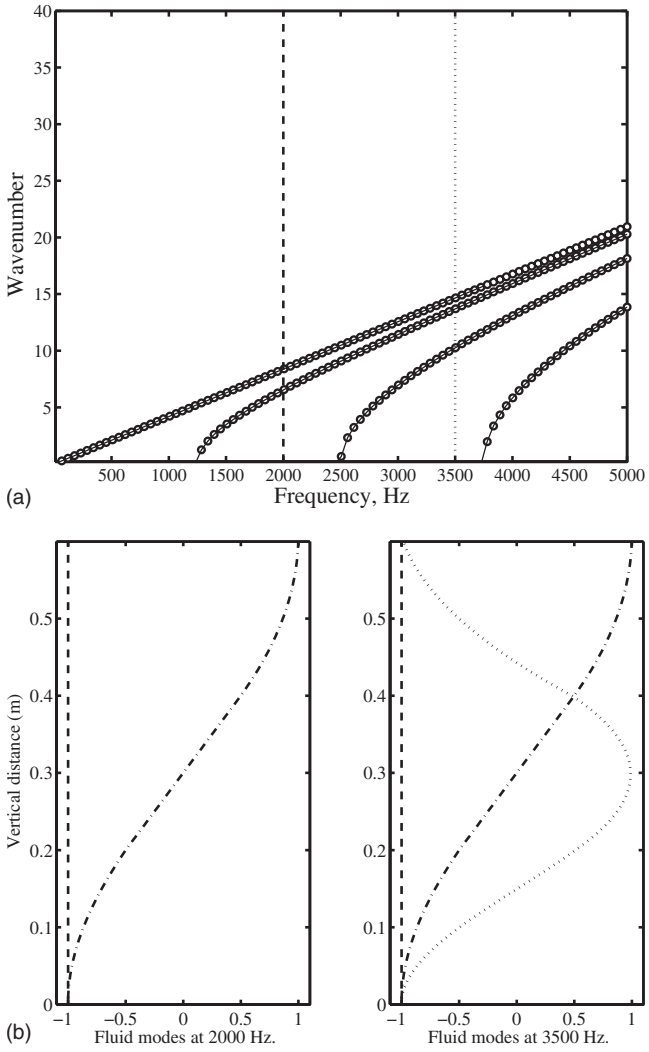
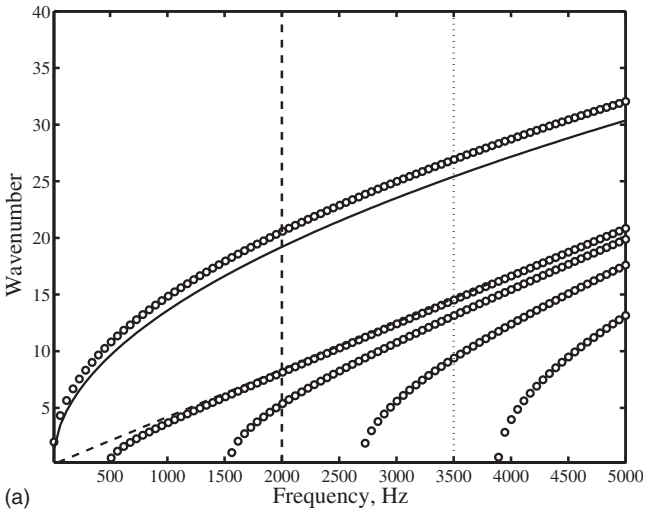


FIG. 2. Dispersion relations depicting real propagating wavenumbers, λ, versus frequency, f, for water-filled rigid duct of height $D=0.6$ m. (a) Circles (○) denote WFE solution curves from Ref. 1 and dashed lines depict exact dispersion curves for fluid. (b) Non-dimensionalized cross-sectional fluid modes Φ, $\|\Phi\|_\infty=1$, at frequencies 2000 and 3500 Hz indicated by vertical lines in (a).

the appearance of propagating waves (cut-on) undergoes a shift to the right eventually reaching the rigid boundary to sound-soft duct cut-on eigenvalues. Specifically as the plate thickness decreases Figs. 3(a) and 4(a) show the propagating plane-wave mode curve tending toward the zero pressure value cut-on eigenvalue, corresponding to 625 Hz. The mode shape is shown by dashed line in Figs. 3(a) and 4(a). Similarly the first and second order modal curves also shift to the right along the frequency axis.

On comparing Figs. 3(a) and 4(a) to Fig. 2(a) an upper branch showing the advent of a structural-like wave is evident. The solid lines in Figs. 3(a) and 4(a) show the uncoupled structural curve in the vicinity of this coupled dispersion curve. The corresponding modeshapes for these curves are given by solid lines in Figs. 3(b) and 4(b). Generally the coupled dispersion curves remain close to the uncoupled curves and undergo transitions between the *in vacuo* structural and the uncoupled acoustic solutions. The mode shape, given by a solid line, clearly has a *near-field*-like



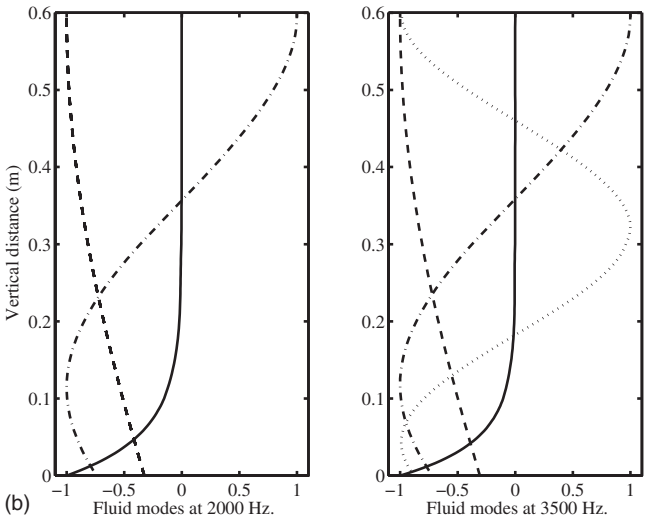
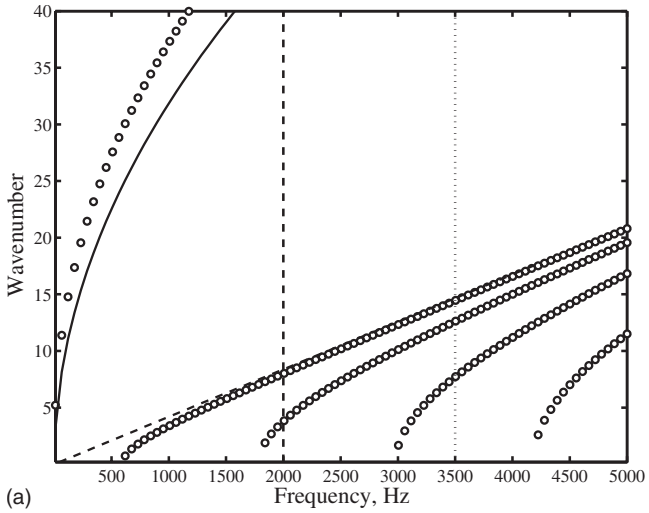


FIG. 3. Dispersion relations, similar to Fig. 2, showing real propagating wavenumbers versus frequency for water-filled steel-backed duct of height $D=0.6$ m with thickness of steel plate, $t=0.022$ m. (a) Circles (○) denote WFE solution curves from Eq. (19). Solid lines correspond to *in vacuo* dispersion relations. (b) Non-dimensionalized cross-sectional fluid modes Φ, $\|\Phi\|_\infty=1$, at frequencies 2000 and 3500 Hz indicated by vertical lines in (a).



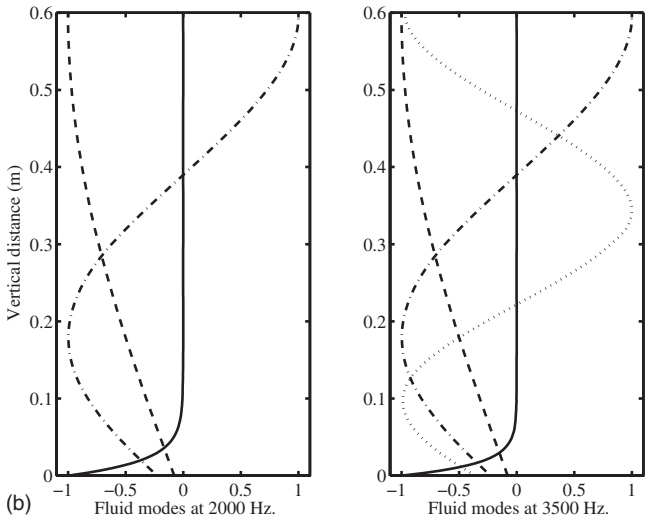


FIG. 4. Dispersion relations showing real propagating wavenumber λ versus frequency for water-filled steel-backed duct of height $D=0.6$ m with thickness of steel plate, $t=0.004$ m. (a) Circles (○) denote WFE solution curves from Eq. (19). Lines correspond to *in vacuo* dispersion relations. (b) Non-dimensionalized cross-sectional fluid modes Φ, $\|\Phi\|_\infty=1$, at frequencies 2000 and 3500 Hz indicated by vertical lines in (a).

characteristic. As the plate thickness decreases it is shown in Figs. 3(b) and 4(b) that this near-field feature is more prominent. It is also evident that the corresponding fluid modes, from a cross-section with rigid sides, Fig. 2 case, also undergo significant variations as the plate thickness decreases, Figs. 3 and 4.

An important observation here is that the SFEM should be able to capture all the propagating waves in the vibro-acoustic system. Otherwise the physics of the original system and the SFEM semi-analytical model will not be compatible. Hence the number of degrees of freedom, and hence the size of the model, should be sufficiently large so that all the cross-section propagating mode shapes are sufficiently approximated. Although this approximation is essentially one-dimensional the size of the model is clearly frequency dependent and warrants a two-dimensional error analysis. Numerical errors for simple two-dimensional acoustic problems have been discussed in Ref. 1.

B. Validation example for open-ended box

Section V A considers the free propagating modes, as a subset of the WFE derivation. However, we are primarily concerned with fluid-structure interaction, and therefore we now proceed to apply SFEM to the case of a fluid volume bounded by a plate that is excited by mechanical forces. In order to illustrate the high accuracy that can be obtained with the SFEM, the case is considered of a steel plate of length $L_P=0.5$ and thickness $t=0.001$ m backed by a water-filled cavity of depth $D=0.6$ m, see Fig. 5, where the right-hand panel undergoes anechoic termination. Simulations are compared against FEMLAB multiphysics calculations.[17] The triangular mesh for cubic shape functions for the FEMLAB calculations is shown in Fig. 5. To compare the performances of the SFEM with the existing element-based techniques, coupled finite-element multiphysics models of the considered problem have been solved with FEMLAB 3.1.[17] The structural FEM meshes consist of triangular isotropic plane strain two-dimensional elements; similarly the acoustic FEM mesh

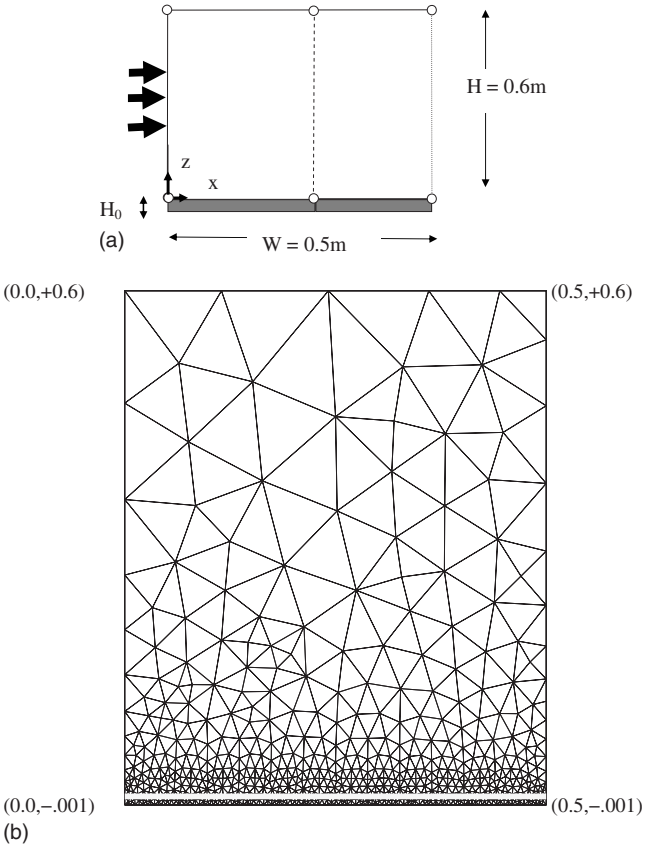


FIG. 5. Geometries for SFEM and FEMLAB methods. Excitation is a unit normal acceleration denoted by arrows with an anechoic termination boundary on the right-hand wall: (a) spectral finite-element "mesh" for the coupled vibro-acoustic problem; (b) illustration of a standard triangulated mesh for coupled problem showing a 1312 triangular graded mesh close to the fluid-structure interface where structure plate thickness equals $t=0.001$ m.

consists of triangular fluid elements. Cubic interlopatory shape functions are used throughout. The steps toward generating a suitable mesh are not described here but the mesh for the thin elastic solid region is quite fine due to the geometry. The white line in Fig. 5(b) shows the interface between solid and fluid regions, and normal acceleration is assumed continuous between the fluid and structural regions. To compare with FEMLAB simulations the plate edges are chosen simply supported equation (3) in the examples shown here and the dynamic stiffness matrix is derived using wave functions given in Eq. (23).

Figure 6 shows the normal plate displacement predictions and the normal fluid pressure predictions along the fluid-plate structure interface at 830 Hz. Solutions obtained with the SFEM wave model consist of 57 degrees of freedom with two SFEM elements each derived by local dynamic 19×19 stiffness matrix. Calculations were performed against a FEMLAB multiphysics simulation using 1312 cubic shape functions defined over triangular elements for the structural and acoustic shape functions totaling 1808 degrees of freedom. For computational expenditure for single frequency calculations, FEMLAB totaled 4.5 s in FEMLAB to complete pre- and post-processing simulations to illustrate solutions, Fig. 6. CPU expenditure time for wave-based SFEM simulation was 0.18 s for solving and post-processing the solution. Assuming the exact solution is attained by FEMLAB software relative errors for the SFEM technique were less than 2.0% on average as shown in Fig. 6.

C. Transmission loss for flexible muffler

To predict the acoustic performance of a complex pipe system a plane wave acoustic source is assumed. This is achieved through a given uniform mechanical velocity source and is included in the calculation of the transmission loss (TL) of the system.[18] The derivation of TL, as well as their use with the numerical methods, is based on plane wave propagation sound sources (i.e., the entire face of the inlet section moving in unison) and an anechoic termination, i.e., $\zeta=\rho_0 c_0$ at the left- and right-hand end sections of the silencer region, see Fig. 7. This models completely anechoic source and termination conditions at inlet and outlet and is easily implemented in the geometrical definitions for SFEM and FEMLAB models. The dimension of the regions used in the FEMLAB and SFEM calculations are illustrated in Fig. 7. The geometry of each model is identical with the velocity source assumed to be situated at the inlet pipe, $x=-0.5$ m at the left-hand side. Note that the length of the outlet pipe for the SFEM may have had arbitrary axial length but was fixed to the FEMLAB geometry for comparison purposes. The lengths of the inlet and outlet pipes were 0.5 m in both simulation models. Nonetheless TL results shown in Fig. 9 were found to be identical for all arbitrary length outlet pipes considered in the SFEM model.

Firstly, for validation, Fig. 8 shows a comparison between FEMLAB and SFEM calculations for the geometry and graded mesh close to plate of thickness $t=H_0=0.001$ m shown in Fig. 7 for a given uniform unit velocity at the left-hand panel. The computational expenditure for this single frequency calculation is similar to the example in Sec. V B. The FEMLAB simulation consisted of 2732 DOFs against 58 DOFs for the SFEM for the complete assembly. The depth of the chamber was fixed at $D=0.05$ m and height of the inlet and outlet pipes $H=0.13$ m. Assuming accuracy of the FEMLAB solution the averaged relative errors for plate displacement and acoustic pressure at the interface, shown in Figs. 8(b) and 8(c), were 2.70% and 3.43%, respectively.

Following this validation the TL of the complex silencer system is now shown. The definition of TL is the ratio of the incident sound power to the transmitted sound power. As long as the inlet and outlet regions of the muffler are of the same cross-section, and the properties of the fluid density do not change, then the TL can be expressed as

$$\mathrm{TL}=20\log 10\left|\frac{P_i}{P_t}\right|, \tag{27}$$

where P_i is the root mean square (rms) pressure of the incident wave, and P_t is the rms pressure of the transmitted wave with chamber in place. Assuming a plane wave propagating source P_i and P_t are easily calculated; see Ref. 18. This can be simplified to the following equation: $\mathrm{TL}=\mathrm{SPL}_i-\mathrm{SPL}_t$ where it is understood that SPL_i is obtained without the silencer in place, and SPL_t is obtained with the silencer in place. Hence TL shows loss of energy expected through the waveguide system show in Fig. 7. The solutions presented in

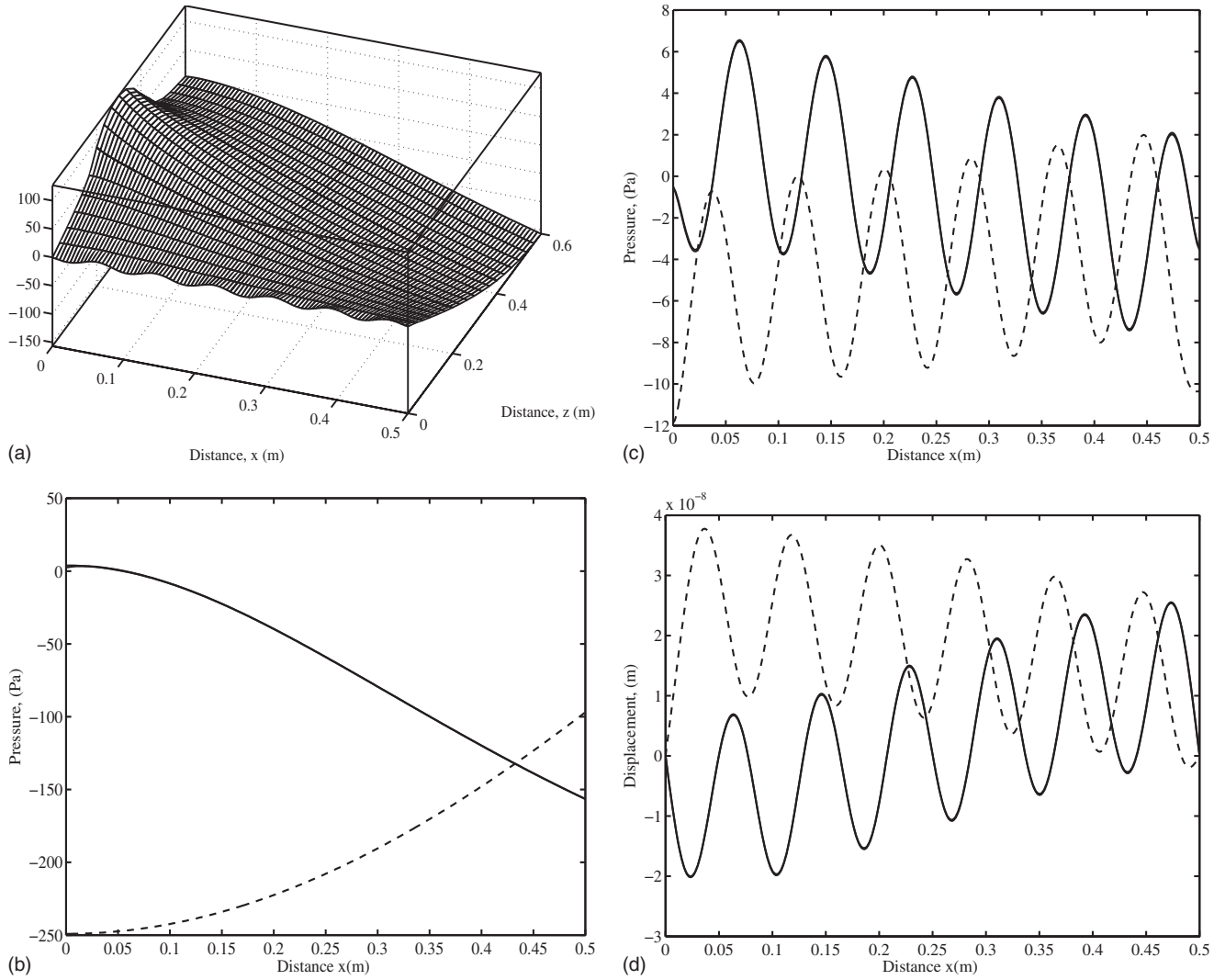


FIG. 6. Solutions for coupled vibro-acoustic problem, water-filled rigid box enclosed by steel plate, thickness $t=0.001$ m at frequency 830 Hz. Comparisons with FEMLAB [1808 degrees of freedom (DOFs)] and SFEM (57 DOFs): (a) a mesh plot for real part of pressure and remaining plots; solid line: real; dashed line: imaginary; (b) acoustic pressure at $z=0.6$, relative error 0.40%; (c) acoustic pressure at $z=0.0$, relative error 2.66%; and (d) displacement of plate at $z=0.0$, relative error 1.91%.

Fig. 9 are attainable in FEMLAB taking 10 CPU minutes and SFEM less than 1 CPU minute for 290 calculations. Note that no significant post-processing expenditure was used here. Figure 9 shows results from SFEM and FEMLAB simulations in a logarithmic frequency range from 10.0 to 3000.0 Hz in 290 steps. The TL for a fluid-only silencer combination is seen in Fig. 9(a) where the depth of the chamber is increased from $D=0.01$, 0.05, 0.09 m to its maximum value $D=0.13$ m. The relative change in TL is about 1.0 dB between each case, the maximum TL attained for the $D=0.13$ m case. Note that plane wave transmission is no longer possible beyond 2900 Hz, evident by the presence of the peak in the solid line in Fig. 9(a).

Next follows the introduction of the plate structure as shown in Fig. 7(b). The system is the same as illustrated in Fig. 9(a) and curves denoted equivalently, except that in this case the chamber is backed by a fixed given plate with thickness $t=0.004$ m. A striking difference in the TL of the system is clearly observable in Fig. 9(b) due to the introduction of a coupled element in the water-filled chamber. Evidence of the complex energy exchange between the fluid and the structure is shown even in the shallowest silencer chamber $D=0.01$ m given by dotted curve in Fig. 9(b). This plot shows significant TL up to 200 Hz with a slight pass-band between 200 and 550 Hz.

Figure 9(c) demonstrates the influence of the mass and bending stiffness of the plate for fixed chamber depth $D=0.01$ m. Again the objective of a muffler is that TL should be small for low frequencies, so that steady flow is not inhibited and TL is higher at high frequencies to inhibit vibro-acoustic noise. We see from this plot that where the "mass" of the plate is lower the silencer behaves like a high-pass filter. However, the silencer should act as a low-pass filter not inhibiting low-frequency wave propagation. The system acts as a quarter-wavelength muffler since the fluid modes cut on at finite non-zero frequencies; see Figs. 3 and 4. A trapped-mode arises causing the system to inhibit energy transmission. However, this phenomenon is particular to the plane case of two-dimensional rectangular waveguides. In Ref. 19 this stop-band phenomenon has been studied for

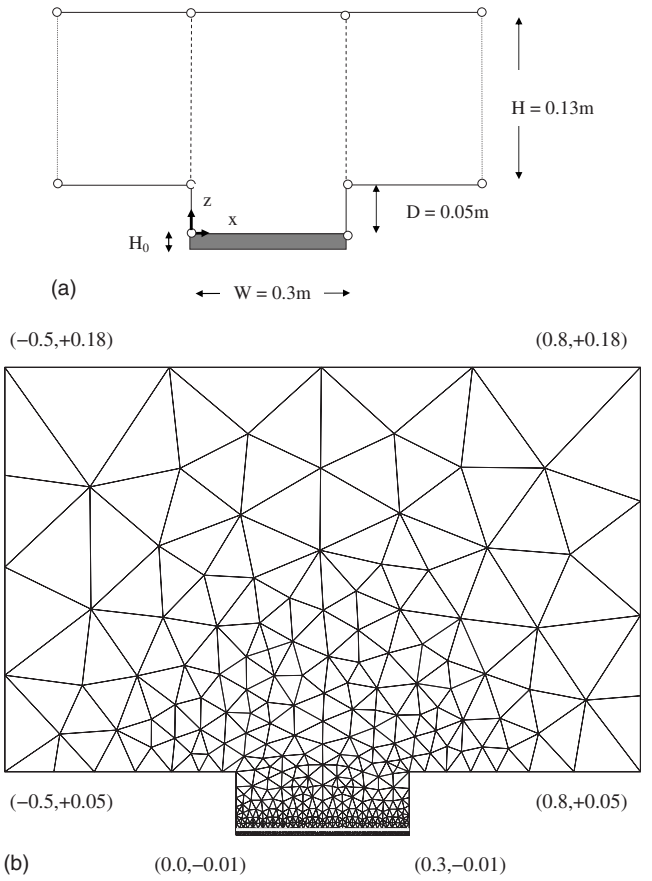


FIG. 7. Geometries for SFEM and FEMLAB methods for flexible muffler example. Excitation is unit normal acceleration at left-hand boundary and anechoic termination boundary on right-hand wall: (a) spectral finite element mesh for coupled vibro-acoustic problem; (b) illustration of standard triangulated mesh for coupled problem showing 1872 triangular elements and graded mesh close to plate of thickness $t=H_0=0.001$ m.

elastic inclusions in fluid-filled cylinders recently. Here, though, a fluid mode exists at arbitrary small frequencies. Thus the trapped-mode phenomena cannot occur at very low frequencies.

VI. CONCLUSIONS

A spectral method in the form of a finite-element scheme has been used to treat the problem of sound transmission in uniform and non-uniform waveguides carrying an acoustic fluid backed by a structural element. A unique feature of the spectral finite-element approach is the use of basis functions generated from linear generalized eigenvalue calculations. The basis functions themselves, which are solutions to the homogeneous reduced wave equations, may be defined over regions of arbitrary length backed by structural elements. Direct wavenumber comparison with a standard FEM solver has been shown to yield satisfactory results for fluid-structure interaction examples. In particular, it has been shown that the TL and hence the energy exchange of a simple silencer assembly, with a heavy fluid, are sensitive to the structural interact ion with the fluid for deep as well as very shallow chambers.

The waveguide geometries that have been given the most attention in the present study are that of a rectangular duct lined by a structure on the side of a geometrical non-

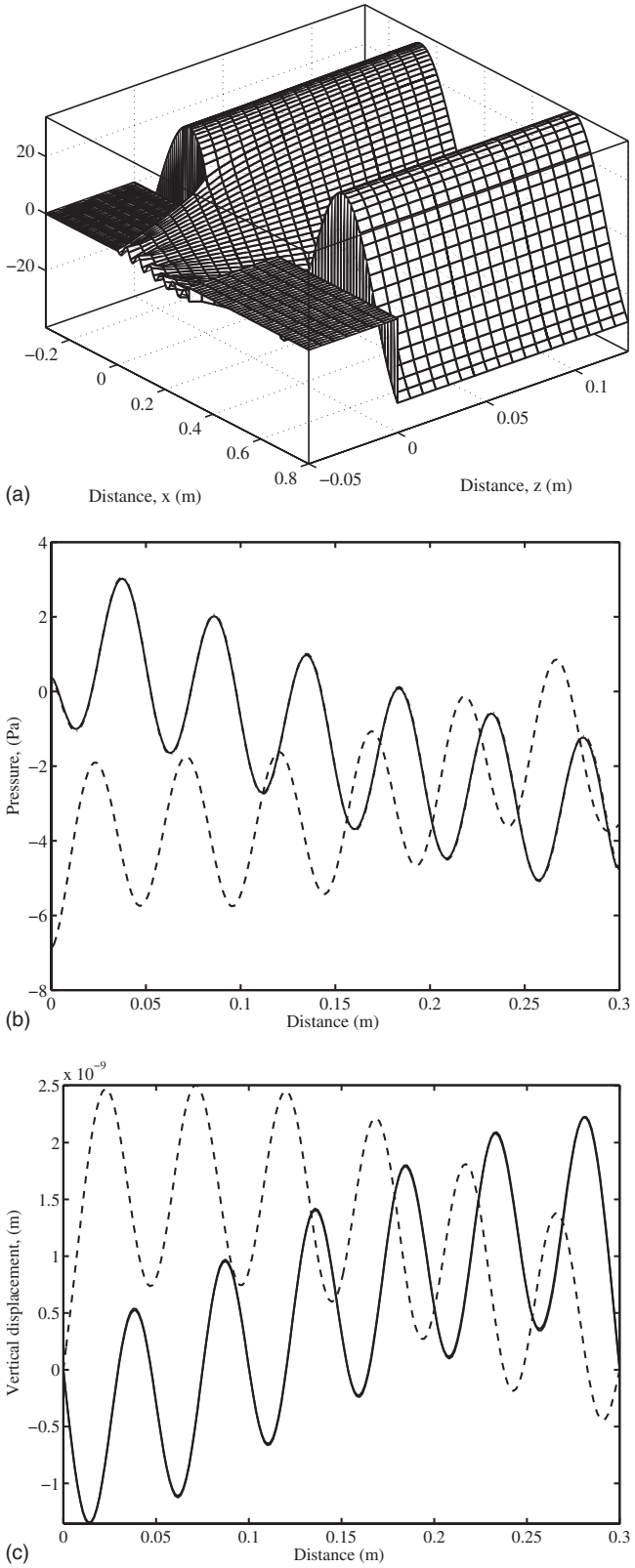
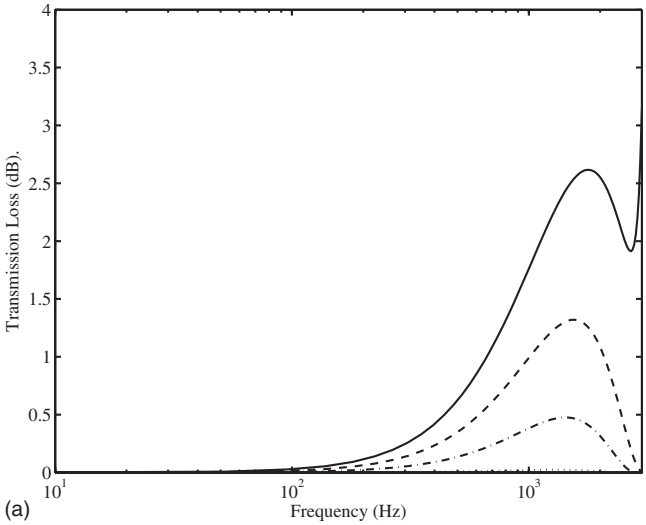
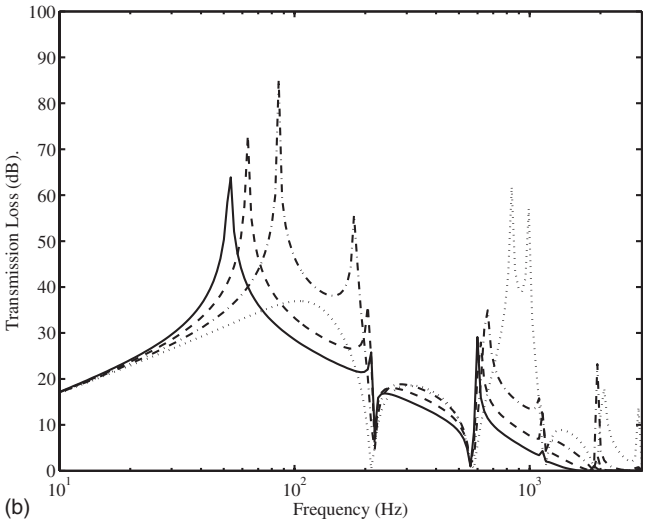


FIG. 8. Solutions for coupled vibro-acoustic problem, water-filled silencer geometry shown in Fig. 7 enclosed by steel plate, thickness $t=0.001$ m at 2730 Hz. Comparisons with FEMLAB (2732 DOFs) and SFEM (58 DOFs): (a) mesh plot for real part of pressure and remaining plots; solid line: real; dashed line: imaginary; (b) acoustic pressure at $z=0.0$, relative error 3.43%; and (c) displacement of plate at $z=0.0$, relative error 2.70%.

uniform rectangular duct. These configurations are not easily amenable to analytical treatment and the use of a numerical method is appropriate here. The finite-element formulation in





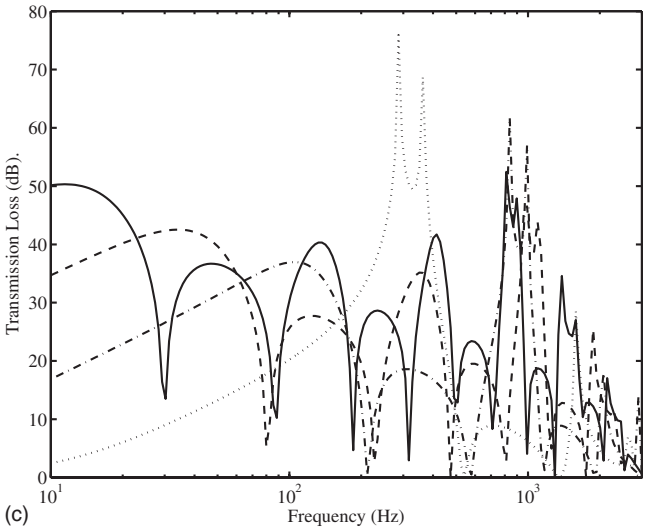


FIG. 9. TL versus frequency for geometry design in Fig. 7. Notation for (a)–(c) in order (i) dots, (ii) dashed-dotted, (iii) dashed, and (iv) solid line (a) Water-filled rigid ducts, various heights of chamber, $D=0.01$ (i), 0.05 (ii), 0.09 (iii), and 0.13 m (iv); (b) Water-filled flexible muffler, varying chamber heights as in (a) with steel plate thickness $t=0.004$ m; (c) water-filled flexible muffler, fixed chamber height, $D=0.01$ m, and varying plate thicknesses, $t=0.01$, 0.004, 0.002, 0.001 m.

this investigation is, however, sufficiently general that it can be extended to any non-uniform rectangular waveguide with varying linings, fluid densities, and structural boundaries and may be coupled with a standard finite-element code for the study of more complicated geometries.

ACKNOWLEDGMENTS

The author wishes to acknowledge support from an EU FP6 Marie Curie Individual Fellowship (Grant No. HPMF-CT-2000-01080) and Svante Finnveden for many of the subroutines and technical discussions necessary to complete this work.

[1]A. T. Peplow and S. Finnveden, "A super-spectral finite element method for sound transmission in waveguides," J. Acoust. Soc. Am. **116**, 1389–1400 (2004).
[2]C. Song, "Dynamic analysis of unbounded domains by a reduced set of base functions," Comput. Methods Appl. Mech. Eng. **194**, 4075–4094 (2006).
[3]I. Babuška and J. M. Melenk, "The partition of unity method," Int. J. Numer. Methods Eng. **40**, 727–758 (1997).
[4]O. Lagrouche and P. Bettess, "Short wave modelling using special finite elements—towards and adaptive approach," in *The Mathematics of Finite Elements and Applications, X*, edited by J. Whiteman (Elsevier Science, Oxford, 2000), pp. 181–194.
[5]S. Finnveden, "Exact spectral finite element analysis of stationary vibrations in a railway car structure," Acta Acust. **2**, 461–482 (1994).
[6]S. Finnveden, "Spectral finite element analysis of the vibration of straight fluid-filled pipes with flanges," J. Sound Vib. **199**, 125–154 (1997).
[7]C.-M. Nilsson and S. Finnveden, "Waves in thin-walled fluid-filled ducts with arbitrary cross-sections," J. Sound Vib. **310**, 58–76 (2007).
[8]F. Fahy, *Sound and Structural Vibration: Radiation, Transmission and Response* (Academic, London, 1987).
[9]C. R. Fuller and F. J. Fahy, "Characteristics of wave propagation and energy distributions in cylindrical elastic shells filled with fluid," J. Sound Vib. **81**, 501–518 (1982).
[10]G. Pavic, "Vibroacoustical energy flow through straight pipes," J. Sound Vib. **142**, 293–310 (1990).
[11]J. B. Lawrie and I. M. M. Guled, "On tuning a reactive silencer by varying the position of an internal membrane," J. Acoust. Soc. Am. **120**, 780–790 (2006).
[12]C. Wang, J. Han, and L. Huang, "Optimization of a clamped plate silencer," J. Acoust. Soc. Am. **121**, 949–960 (2007).
[13]L. Huang, "Broadband sound reflection by plates covering side-branch cavities in a duct," J. Acoust. Soc. Am. **119**, 2628–2638 (2006).
[14]L. Huang and Y. S. Choy, "Vibroacoustics of three-dimensional drum silencer," J. Acoust. Soc. Am. **118**, 2313–2320 (2005).
[15]Y. S. Choy and L. Huang, "Experimental studies of a drumlike silencer," J. Acoust. Soc. Am. **112**, 2026–2035 (2002).
[16]F. Tisseur and K. Meerbergen, "The quadratic eigenvalue problem," SIAM Rev. **43**, 235–286 (2001).
[17]FEMLAB, COMSOL Users Manual, 2004.
[18]S. Bilawchuk and K. R. Fyfe, "Comparison and implementation of the various numerical methods used for calculating transmission loss in silencer systems," Appl. Acoust. **64**, 903–916 (2003).
[19]S. V. Sorokin and O. A. Ershova, "Plane wave propagation and frequency band gaps in periodic plates and cylindrical shells with and without heavy fluid loading," J. Sound Vib. **278**, 501–526 (2004).

A correction of random incidence absorption coefficients for the angular distribution of acoustic energy under measurement conditions

Cheol-Ho Jeong[a)]
Department of Electrical Engineering, Acoustic Technology, Technical University of Denmark (DTU), DK-2800 Kgs. Lyngby, Denmark

(Received 18 June 2008; revised 24 November 2008; accepted 19 January 2009)

Most acoustic measurements are based on an assumption of ideal conditions. One such ideal condition is a diffuse and reverberant field. In practice, a perfectly diffuse sound field cannot be achieved in a reverberation chamber. Uneven incident energy density under measurement conditions can cause discrepancies between the measured value and the theoretical random incidence absorption coefficient. Therefore the angular distribution of the incident acoustic energy onto an absorber sample should be taken into account. The angular distribution of the incident energy density was simulated using the beam tracing method for various room shapes and source positions. The averaged angular distribution is found to be similar to a Gaussian distribution. As a result, an angle-weighted absorption coefficient was proposed by considering the angular energy distribution to improve the agreement between the theoretical absorption coefficient and the reverberation room measurement. The angle-weighted absorption coefficient, together with the size correction, agrees satisfactorily with the measured absorption data by the reverberation chamber method. At high frequencies and for large samples, the averaged weighting corresponds well with the measurement, whereas at low frequencies and for small panels, the relatively flat distribution agrees better.
 [DOI: 10.1121/1.3081392]

PACS number(s): 43.55.Ev, 43.55.Nd [LMW] Pages: 2064–2071

I. INTRODUCTION

A diffuse field is an idealized concept, and it is the basis of several standardized measurements, e.g., the absorption of materials, the sound power of noise sources, and the transmission loss of walls. For a perfectly diffuse field, two assumptions must be satisfied (for example, see Ref. 1).

(1) The local energy density in a room is uniform (the spatial diffusion).
(2) The energy is uniformly incident onto a surface from all directions (the directional diffusion).

In a reverberation room except regions near boundaries, the first condition may be satisfied. At locations about one wavelength apart from boundaries, the deviation of sound pressure is bounded within 1 dB in a large reverberation chamber (0.7λ from corners and edges and 0.25λ from surfaces).[2] Despite attempts to obtain a diffuse sound field in a reverberation room, the second assumption of the random incidence is very hard to fulfill. In particular, when a test specimen covers one surface as in the absorption and sound transmission loss measurement, an ideal diffusivity is rarely obtained. As a compensation, truncation of the angle of incidence was introduced in the calculation of transmission loss, and typical values of the limiting angle vary from 70° to 85° based on empirical data.[3,4] This implies that acoustic energies are not uniformly incident on a sample surface. From this observation, one may deduce that grazingly incident energies are negligible.

The random incidence absorption coefficient is based on the assumption that the intensities of the incident sound are uniformly distributed over all possible directions. In addition, a random distribution of phase of the wave incident on the wall is also assumed. These assumptions result in Paris formula associated with $\sin(2\theta)$.[5] However, the measured absorption coefficient by Sabine's formula (sometimes referred to as the Sabine absorption coefficient) shows a significant discrepancy compared with the random incident coefficient from measured impedance data.[6] It is known that measured absorption coefficients are overestimated for small absorber samples and sometimes exceed unity, even for a nearly locally reacting surface. Theoretically, the random incidence absorption coefficient cannot exceed 0.951 for the specific surface impedance of 1.567. Reasons for the overestimation have been widely accepted to be the finite size of samples and the edge diffraction.[7–12]

To match the theoretical values with measured data, several works have been studied. Concerning the sound transmission loss, de Bruijn[13] addressed the influence of the degree of diffusivity in measurement conditions. Kang *et al.*[14] employed several Gaussian weighting functions for the calculation of sound transmission loss to achieve better agreements with the measurement. Kang *et al.*[15] also measured the incident energy density onto a sample in a reverberation chamber using sound intensimetry. The measured incident intensity decreases with increasing incidence angle; however,

a)Author to whom correspondence should be addressed. Electronic mail: chj@elektro.dtu.dk

 0001-4966/2009/125(4)/2064/8/$25.00

it becomes rather uniform at low frequencies. For the absorption measurement, Makita and Fujiwara[16] investigated possible maximum and minimum values of absorption coefficients assuming non-uniform incident energy, although they had no idea of the angular distribution of the incident energy distribution. They tested unrealistic weighting functions, but this research revealed that the absorption coefficient can exceed unity depending on the directional energy distribution. Makita and Hidaka[17] suggested a revised cosine law of oblique incident energy considering the effective mean free paths of sound rays and reverberation time. However, this modified function is not experimentally validated. Thomasson[18,19] concentrated mainly on the size effect of an absorbing sample, by introducing the concept of averaged radiation impedance. The averaged radiation impedance varies with panel size and aspect ratio under the assumption that the radiation impedance depends weakly on the azimuth angle. However, an ideal diffuse field was assumed in his study. Several semi-empirical corrections have also been suggested. London[20] defined a quantity called the equivalent real impedance and used this quantity for computing the reverberation absorption coefficient. Atal[21] suggested an empirical correction by multiplying a frequency-dependent complex quantity in the form of $(a+bj)$. Dämmig[22] examined the effect of sample locations and room shapes on the sound absorption using the ray tracing method. This work is based on statistical analyses of incidence angles, not the incident energies, to explain the variation of measured absorption coefficients in several source locations and rooms.

In the present study, angular distributions of energy density incident onto an absorbing sample have been investigated using the beam tracing method. The final objective of this study is to improve the agreement by weighting the calculated absorption coefficient according to the actual incident energy density instead of assuming an ideally diffuse incident field.

II. ANGULAR DISTRIBUTION OF ENERGY DENSITY AS A WEIGHTING FUNCTION

First of all, the angular distribution of the reverberant energy density must be calculated. In this study, the beam tracing method[23–25] is employed to simulate the sound propagation in a reverberation chamber. Triangular beams, which are emitted from a source, are followed by their central axes (hereafter rays) without splitting algorithm.[26,27] Instead of splitting beams when beams intersect more than one surface at edges or corners, a large number of beams having a sufficiently small solid angle emanate and illuminate a reverberant field. Only specular reflection is adopted during the tracing, and a negligible absorption coefficient of 0.01 is assigned for all surfaces of a reverberation chamber irrespective of the frequency. The incidence angles and energies coming from these angles, which are computed by the geometrical acoustics technique, will eventually construct an angular distribution of incident energy density as a function of the incidence angle.

Two chambers were chosen as test examples: one is a rectangular parallelepiped chamber and the other is a room

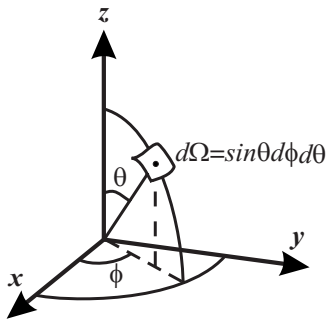


FIG. 1. Spherical polar coordinate representing solid angle element.

with non-parallel surfaces. In accordance with ISO 354 (Ref. 28) and ASTM C 423–02,[29] the sources are located at trihedral corners of the rooms. The assumption that an absorber sample covers one whole surface of the room facilitates the simulation in a way that one can collect all the information of beams incident onto the specific surface. This surface, on which an absorber is installed, is called a target surface.

During the beam tracing, the information on the directional energies and angles of incidence is saved. Acoustic energy decays inversely proportional to r^2 and it is reduced by $(1-\alpha_i)$ whenever beam hits surfaces. Here, r denotes the traveling distance and α_i denotes the absorption coefficient of the ith surface. For a steady-state condition, a total directional energy from θ_i is calculated by summing all components over an interval, $[\theta_{i,l}\ \theta_{i,u}]$. Here, $\theta_{i,l}$ and $\theta_{i,u}$ are the lower and upper limits of the interval, respectively, and θ_i is the arithmetic mean of two limiting values. The directional energy density is the ratio between the total incident sound energy, E_θ, and the corresponding solid angle, Ω_θ. Using a generalized concept of solid angle in terms of spherical polar coordinates, θ and ϕ, respectively (see Fig. 1), the solid angle element is expressed as $d\Omega=\sin\theta d\phi d\theta$. By integrating the azimuth angle, ϕ, from 0 to 2π and the polar angle, θ, over the corresponding interval of $[\theta_{i,l}\ \theta_{i,u}]$, one can find the solid angle at θ_i, as follows:

$$\Omega_{\theta_i}=\int_0^{2\pi}\int_{\theta_{i,l}}^{\theta_{i,u}} d\Omega=\int_0^{2\pi}\int_{\theta_{i,l}}^{\theta_{i,u}} \sin\theta d\theta d\phi$$
$$=2\pi\int_{\theta_{i,l}}^{\theta_{i,u}} \sin\theta d\theta. \quad (1)$$

The solid angle in Eq. (1) increases with the angle of incidence. The directional energy density is defined as the total directional energy incident over the interval of $[\theta_{i,l}\ \theta_{i,u}]$ divided by the corresponding solid angle as follows:

$$D(\theta_i)=\frac{E_{\theta_i}}{\Omega_{\theta_i}}. \quad (2)$$

When the reverberant energy density is computed, the result will be utilized as a weighting function in the calculation of absorption coefficient. In Sec. V, the effect of weighting functions will be discussed by comparing the angle-weighted absorption coefficients with the measured absorption data in a reverberation chamber.

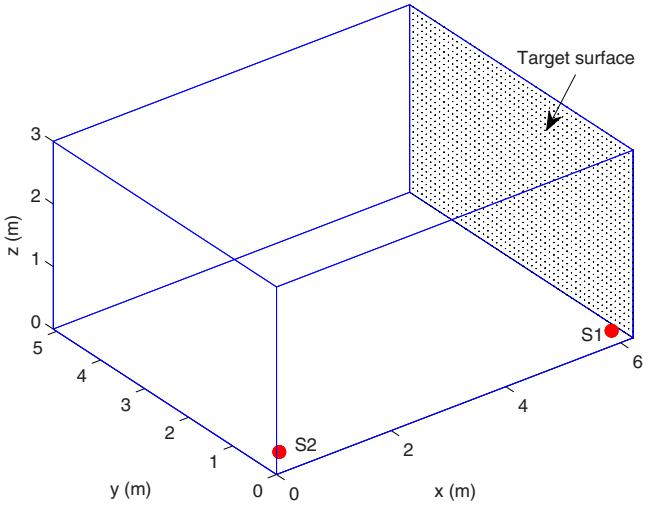


FIG. 2. (Color online) A rectangular room. Target surface is (6.2 m, y, z) plane and two corner sources are located at (6 m, 0.2 m, 0.1 m) and (0.2 m, 0.2 m, 0.2 m), respectively.

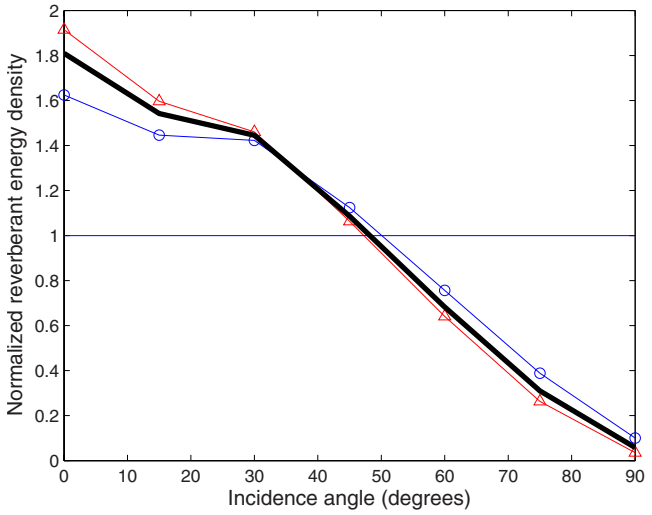


FIG. 3. (Color online) A comparison of reverberant energy density for two corner sources. (—○—) S1 at (6 m, 0.2 m, 0.1 m), (—△—) S2 at (0.2 m, 0.2 m, 0.2 m), and (thick solid line) averaged distribution over 90 source locations.

III. A RECTANGULAR ROOM

In general, lecture rooms and some reverberation chambers are rectangular. This kind of room usually has the problem that the interference is strong due to pairs of parallel surfaces. In this study, all surfaces are assumed smooth, and scattering is not taken into account.

A test rectangular room model is shown in Fig. 2. The edge lengths are 6.2 m, 5.1 m, 3.0 m, and the volume is 94.8 m^3. The target surface of interest is parallel to the y-z plane and located at (6.2 m, y, z). In order to examine the effect of source locations, two possible source locations are chosen in accordance with ISO 354[28] and ASTM C 423–02:[29] According to the standards, a loudspeaker should face into the trihedral corners of a room. The first source, S1, is located at (6 m, 0.2 m, 0.1 m) and S2 is located at (0.2 m, 0.2 m, 0.2 m). In the simulation, 8000 beams were used to scan the space. The number of reflections per beam is limited to 50. It goes without saying that directly transmitted energies from the sources to the target surface contribute the most owing to the shortest distance of propagation. (It is not allowed to place a source near a specimen in measuring the sound transmission loss because the direct radiation is too dominant.)

Incident energy densities for two sources are shown in Fig. 3, which are similar to the energy distributions of direct rays, because the directly transmitted energy overwhelms the total acoustic energy. This normalized energy density shows a similarity with the results by Kang and co-workers.[14,15] Because the source S2 is far from the target surface, only a few direct rays (172 out of 8000) strike the surface. It was found that there is an upper limiting angle of direct incidence, 42° for S2 case. That means there is no directly incident intensity above 42°. Accordingly the normally incident energy is higher than the obliquely incident energy. The acoustic energy density from the normal direction (0°) for S2 is 18% higher than that for S1, due to the lack of obliquely incident direct rays. It can be summarized that when a source is located far from the surface, the normally incident energy is more highlighted, whereas more energies from oblique directions are incident onto the surface for source locations close to the surface.

For 90 equally spaced ($6\times5\times3$) source locations, the average angular distribution of reverberant energy density was simulated and displayed in Fig. 3 (thick line). Sources are distributed by combining x values of (1, 2, 3, 4, 5, 6 m) and y values of (1, 2, 3, 4, 5 m) and z values of (0.2, 1.1, 2 m). The averaged reverberant energy density clearly shows that the acoustic energy density above 80° is lower than 0.2, which will be considered negligible.

IV. A ROOM WITH NON-PARALLEL SURFACES

In Fig. 4, a reverberation room with non-parallel surfaces is shown. The volume of the space is 179 m^3 and the

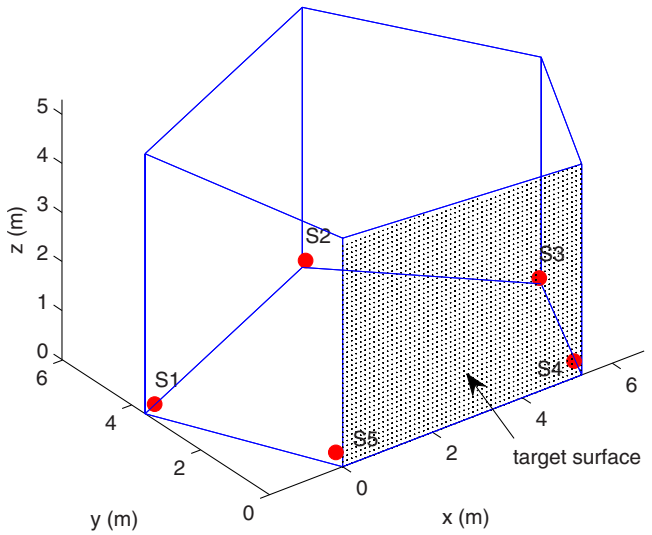


FIG. 4. (Color online) An irregularly shaped room. Target surface is (x, 0, z) plane and five corner sources are located at (−1.5 m, 3.5 m, 0.2 m), (3.7 m, 5.9 m, 0.2 m), (6.6 m, 2.9 m, 0.2 m), (5.3 m, 0.2 m, 0.2 m), and (0 m, 0.2 m, 0.2 m), respectively.

TABLE I. Geometrical nodal data of a reverberation chamber in Fig. 4.

Node numbering	Coordinate		
	x (m)	y (m)	z (m)
1	0	0	0
2	5.32	0	0
3	6.71	2.99	0
4	3.72	6.02	0
5	−1.63	3.61	0
6	0	0	4.66
7	5.32	0	4.30
8	6.71	2.99	4.62
9	3.72	6.02	5.30
10	−1.63	3.61	5.30

absorption coefficient is also 0.01, irrespective of surface and frequency. Geometrical nodal points of the model are listed in Table I. The target surface is x-z plane. Simulations were carried out for five trihedral corner sources shown in Fig. 4.

Figure 5 shows the normalized energy density for five corner sources. For source S1, direct rays from the normal direction cannot reach the target surface. The lowest direct angle of incidence is 23°. By only considering direct sound, the energy density shows a single peak at 45° due to the non-existence of the normal components. However, the reverberant energy density has two peaks at 0° and 45°. This result is apparently different from the rectangular room result, which shows a decreasing tendency with the angle of incidence. The energy is re-distributed due to the lack of direct normal incident rays.

The source S2 is located at the farthest distance of all sources. A high energy density value was found for the normally incident direct ray, as expected. As discussed in Sec. III, a long distance between a source and a target surface emphasizes the importance of the normal direction. An abrupt decrease was found above 45° and the contribution becomes less than 0.5.

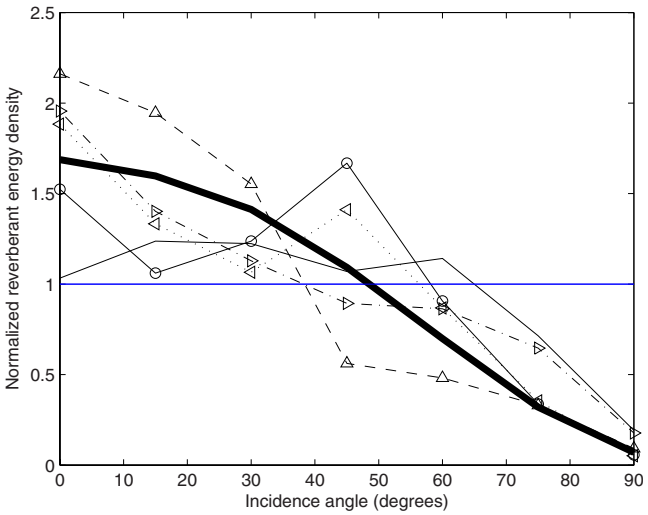


FIG. 5. (Color online) A comparison of angular reverberant energy densities for five corner sources. (—○—) S1, (--△--) S2, (···◁···) S3, (·–▷·–) S4, (—) S5, and (thick solid line) averaged distribution over 96 source locations.

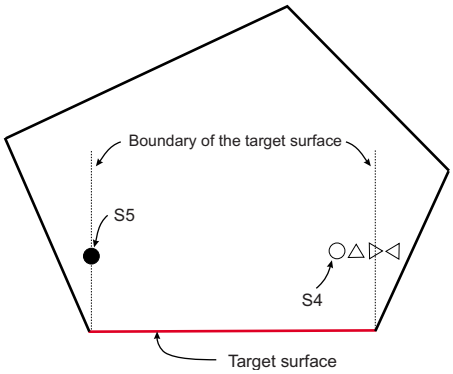


FIG. 6. (Color online) Source locations of S4 and S5 in plane view.

For source S3, the incident energy density shows a strong similarity with that of S1. As the sound source is invisible from the normal direction of the target surface, the incidence angle of 45° becomes pronounced. Therefore a double peak shaped distribution was found for the reverberant acoustic energy density.

The remaining two source locations, S4 and S5, at (5.3 m, 0.2 m, 0.2 m) and (0 m, 0.2 m, 0.2 m), respectively, are situated closer to the target surface. The reverberant acoustic energy densities of two cases show a large difference: The energy continuously decreases for S4, while it is relatively uniform until 60° for S5. In Fig. 6, source locations are illustrated. While the shaded source is located perpendicularly from the boundary of the target surface, two unshaded sources (circle and triangle) are located within the boundary of the target surface. If the source is located within the boundary, the normally incident energy is too pronounced.

Figure 7, which shows the effects of source locations, supports the above statement. By changing x-coordinate of a sound source position from 5.3 to 5.33 m with a step of 0.01 m, one can clearly see that the normal energy density loses its importance, when the source moves away from the surface boundary (refer to the symbols in Fig. 6). When a

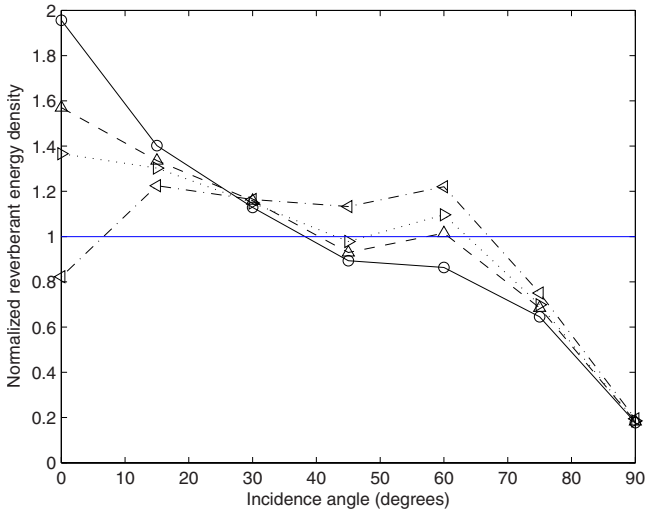


FIG. 7. (Color online) A comparison of reverberant energy densities. x-coordinate of sources varies from 5.30 to 5.33 m with a step of 0.01 m. (—○—) x=5.30 m, (--△--) x=5.31 m, (···▷···) x=5.32 m, and (·–◁·–) x=5.33 m (see Fig. 6).

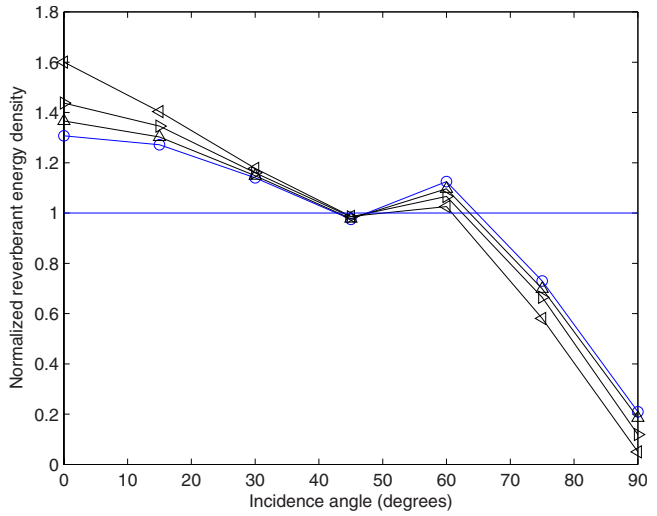


FIG. 8. (Color online) Effects of distance on the reverberant energy densities. (—○—) d=0.01 m, (—△—) d=0.20 m, (—▷—) d=0.50 m, and (—◁—) d=1.00 m. x-coordinate of sources is set to 5.32 m.

source is located outside the boundary, absence of direct normal components decreases overall reverberant energy density for the normal incidence. In order to have a relatively uniform energy distribution, it is desirable to locate the acoustic center of sound source perpendicularly from the surface boundary.

In Fig. 8, effects of a distance between the source and the target surface are shown. Provided that the x-coordinate of source locations is fixed to 5.32 m (perpendicularly from the boundary), distances from the surface to the source were changed to 0.01, 0.2, 0.5, and 1 m. The shorter the distance, the more uniform the reverberant acoustic energy is. It can be concluded that the preferable source location for the uniform reverberant energy is the closest possible and perpendicularly from the boundary of the target surface. For a rectangular room, the ideal position does not exist, because all possible locations are found within the boundary of the target surface.

The thick line in Fig. 5 shows the averaged result over 96 equi-spaced source locations in the irregularly shaped room. The averaged result corresponds well with the rectangular room result, shown in Fig. 3. The calculated reverberant energy densities will be used as a weighting factor in computing absorption coefficients from surface impedances.

V. COMPARISONS WITH THE MEASUREMENT

In Fig. 9, an absorption chart in terms of the real and imaginary parts of the impedance is shown when adopting the averaged energy density for the rectangular room. The solid line and the broken line represent the random incidence absorption coefficient and the angle-weighted absorption coefficient, respectively, calculated as follows:

$$\alpha_{\text{infinite,random}} = \int_0^{\pi/2} \alpha_\theta \sin(2\theta) d\theta, \tag{3a}$$

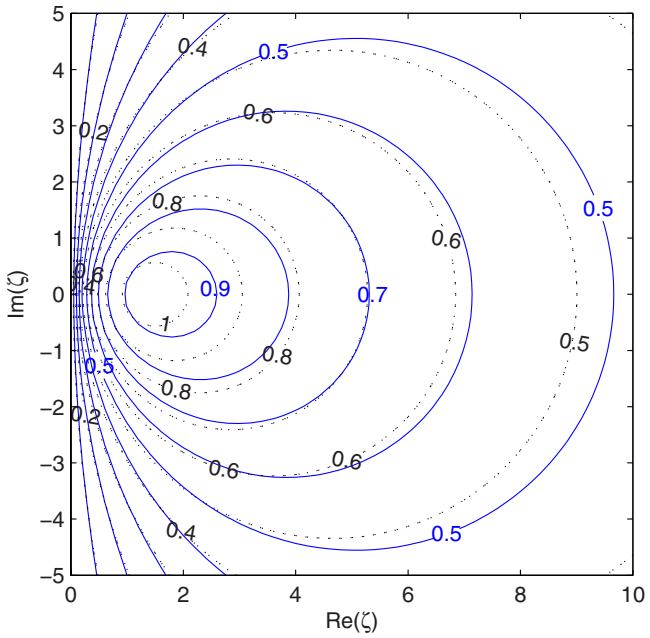


FIG. 9. (Color online) An absorption chart in terms of real and imaginary parts of specific impedance. (——) uniform incidence; (----) angle-weighted.

$$\alpha_{\text{infinite,weighted}} = \int_0^{\pi/2} \alpha_\theta w(\theta) \sin(2\theta) d\theta. \tag{3b}$$

Here, α_θ means the angle dependent absorption coefficient and $w(\theta)$ is the weighting function. According to this chart, if the absorption coefficient is between 0.7 and 0.8, the angle-weighted absorption coefficients are similar to the random incidence absorption coefficients. It is apparent that the maximum random incidence absorption coefficient is less than unity, while the angle-weighted absorption can exceed unity, as in actual reverberation chamber measurements, inside the smallest circle. It should be noted that the result in Fig. 9 cannot be generalized, because the specific weighting function was employed. Moreover, the panel is assumed to be infinitely extended.

It is always necessary to consider the size effect in calculating absorption coefficients. In this study, Thomasson's size correction[18,19] for a finite rectangular panel was adopted. The absorption coefficient for a finite panel is computed using the concept of averaged field impedance as follows:

$$\alpha_{\text{fin}}(\theta) = \frac{4\ \text{Re}(Z_w)}{|Z_w + \bar{Z}_f|^2}. \tag{4}$$

Here, θ is the incidence angle, Z_w is the wall impedance, and $\bar{Z}_f$ is the averaged field impedance over azimuth angle from 0 to 2π expressed as $\bar{Z}_f = 1/2\pi \int_0^{2\pi} Z_f d\phi$. Z_f is the field impedance (sometimes called the radiation impedance) of the wall, which means the impedance of the vibrating surface that radiates sound in the surrounding medium. The field impedance is known as $1/\cos\theta$ for an infinitely large plate; however, the field impedance for a finite sample differs from $1/\cos\theta$, in particular, near grazing incidence.[18,19,30] The field impedance for a finite panel is expressed as follows [see Ref. 18, Eq. (18) and Ref. 19, Eq. (A1)]:

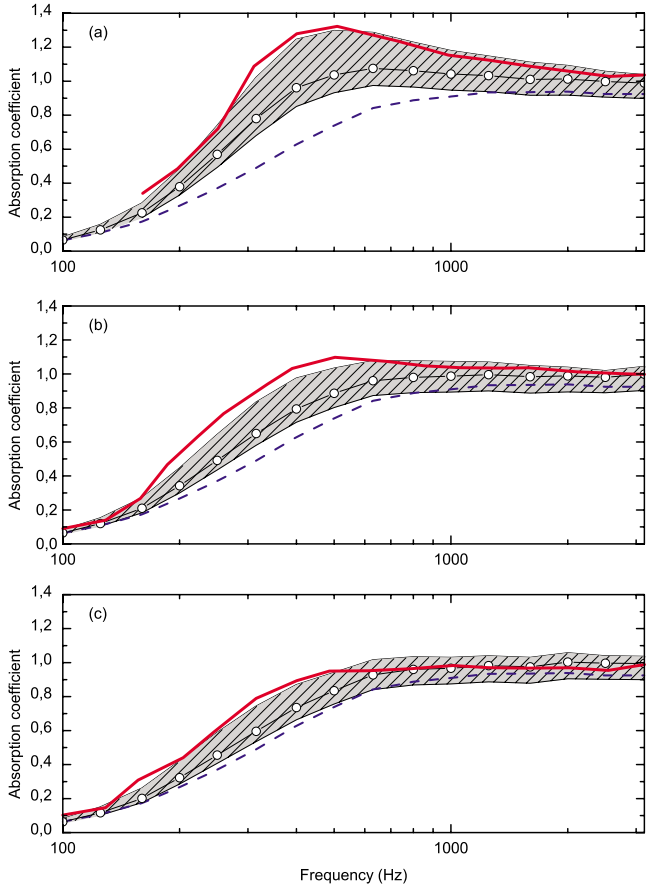


FIG. 10. (Color online) A comparison of absorption coefficient between the reverberation chamber measurement and the angle-weighted value calculated from Z_w. (——) measurement and (—○—) absorption adopting the averaged weighting function. (gray hatched area) Possible range of the angle-weighted absorption coefficient and (– –) random incidence absorption. (a) $e=1.2$ m, (b) $e=2.4$ m, and (c) $e=3.6$ m.

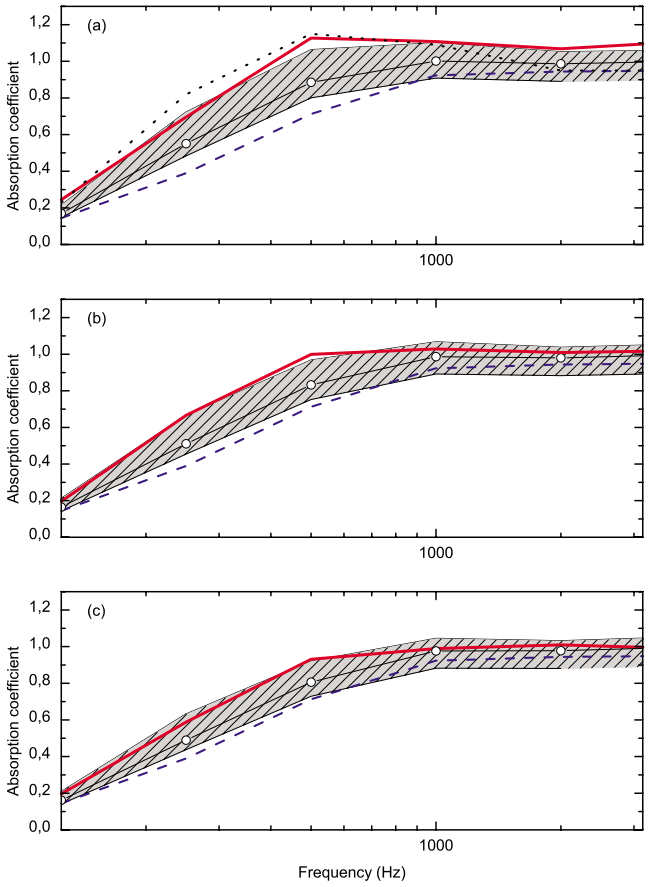


FIG. 11. (Color online) A comparison of absorption coefficient between the reverberation chamber measurement and the angle-weighted value calculated from Z_w. (——) Measurement and (—○—) absorption adopting the averaged weighting function. (gray hatched area) Possible range of the angle-weighted absorption coefficient, (– –) random incidence absorption, and (···) Thomasson's model (Ref. 19). (a) $S=4$ m^2, (b) $S=8$ m^2, and (c) $S=12$ m^2.

$$Z_f = \frac{ik}{S}\int_{S_a}\int\int_{S_a}\int Ge^{ik[\mu_x(x_o-x)+\mu_y(y_o-y)]}dxdydx_ody_o, \quad (5)$$

where k is the wavenumber, $S=\int_{S_a}\int dxdy$, $\mu_x=\sin\theta\cos\phi$, $\mu_y=\sin\theta\sin\phi$, $G=-(2\pi R)^{-1}\exp(ikR)$, and $R=\sqrt{(x-x_o)^2+(y-y_o)^2}$. Consequently the angle-weighted and size-corrected absorption coefficient is calculated as follows:

$$\alpha_{\text{weighted}} = 2\int_0^{\pi/2} w(\theta)\alpha_{\text{fin}}(\theta)\sin(\theta)d\theta$$

$$= 2\int_0^{\pi/2} w(\theta)\frac{4\ \text{Re}(Z_w)}{|Z_w+\bar{Z}_f|^2}\sin(\theta)d\theta, \quad (6)$$

where $w(\theta)$ is the weighting function, which is the angular distribution of energy density by the beam tracing method. The calculated angle-weighted absorption coefficient has been compared with previously published measurements in Figs. 10 and 11. The first absorber consists of 0.05 m thick mineral wool in Thomasson's work.[19] The specific flow resistance and the density of this material were 30 kN s/m^4 and 50 kg/m^3, respectively, and the surface impedance is also taken from Ref. 19. Three kinds of samples were chosen as follows: 1.2×1.2, 2.4×2.4, and 3.6×3.6 m^2. The measured absorption coefficient is the average value from two reverberation chambers with volumes of 190 and 200 m^3, with sufficient diffusion.

The calculated absorption coefficients are shown in Fig. 10 for three different sample sizes. The shaded region denotes a possible range of the angle-weighted absorption coefficient by adopting two limiting angular distributions. The upper curve is obtained by adopting the relatively uniform weighting (for S5 in Fig. 5) and the lower curve is found by adopting the most non-uniform pattern (for S2 in Fig. 5). The hollow circle denotes the angle-weighted absorption coefficient by the averaged weighting function, shown in Fig. 5. Two thick average weighting lines in Figs. 3 and 5 are both similar to a Gaussian distribution. The maximum difference between them is found to be 5% at the normal direction. Because two weighting functions yield nearly the same angle-weighted absorption coefficient, the averaged weighting function in Fig. 5 will be used hereafter. The averaged weighting case can be regarded as a sort of mean value between two limiting absorption coefficients (hereafter, three curves will be named the upper curve, the lower curve, and the averaged weighting, respectively). Large variations in measured data were found depending on sample sizes. In Fig. 10(a), for the smallest sample of edge length (e) of 1.2 m, the maximum value of measured absorption becomes

nearly 1.4. It is apparently due to the size effect, because this kind of overestimation does not occur for the larger two surfaces. One can clearly see that the random incidence absorption from the surface impedance data is far below the measured data. For the smallest sample (a), the measured data correspond well with the upper curve. However, the shaded region becomes narrower, as the frequency increases. When the sample size is four times larger (b), then the measured curve converges to the averaged weighting at high frequencies in Fig. 10(b). Beyond 1 kHz, errors are bounded to 4%. For the largest sample (only this sample is large enough in accordance with ISO 354 specification), the measured curve agrees well with the averaged weighting at frequencies over 800 Hz. One can observe that the measured curve corresponds well with the upper limit value at low frequencies, but the high frequency absorption approaches the averaged weighting case. The three comparisons revealed that the upper curve can be a reasonable guideline at low frequencies.

The second example is the measured data from the second round robin test by Kosten.[11] The material was a rockwool, 5 cm thick, and the density being 100 kg/m^3 (see Ref. 31 for more information, e.g., surface impedance data). Three different surface areas of 4, 8, and 12 m^2 were tested. It should be noted that the measured data are the averaged value over 19 different reverberation chambers all over the world. The degree of diffusion is different from laboratory to laboratory. Dimensions and aspect ratios of samples are not provided. Only the information on sample sizes is given. Samples had different aspect ratios and sometimes different sizes. For example, the surface area varies from 7.5 to 8.25 m^2 for the 8 m^2 sample. Because the author does not have any detailed information on samples, they are regarded as square samples, the edge length e being calculated as $\sqrt{S}$.

Figure 11 shows a comparison between the reverberation chamber measurement and the theoretical angle-weighted absorption coefficient. For the smallest sample in Fig. 11(a), the measured data correspond well with the upper curve, similarly with Fig. 10(a). In this case, Thomasson's model was also compared with the measured data. Thomasson's model[19] seems to agree acceptably below 1 kHz, but it starts to underestimate beyond 1 kHz. In his paper, Thomasson admitted that the deviation at highest frequencies could not be explained. On the other hand, when the proper weighting function is incorporated, the result is improved. It is also observed that the upper curve shows a reasonable guideline at low frequencies and for small size, whereas the averaged weighting agrees well with the measurement for high frequencies and larger samples.

Therefore, a non-dimensionalized parameter, ke, which is multiplication of the wavenumber and the characteristic length of a sample, is introduced to effectively indicate the general trend of the relative errors. A high value of ke means a high frequency and/or a large sample. For small ke, one can see that the upper curves agree well with the measurement. As ke increases, the averaged weighting curve becomes closer to the measurement. The lower curves seem meaningless for all cases, because most of measured absorptions are higher than the angle-weighted absorption coefficient by the averaged weighting function.

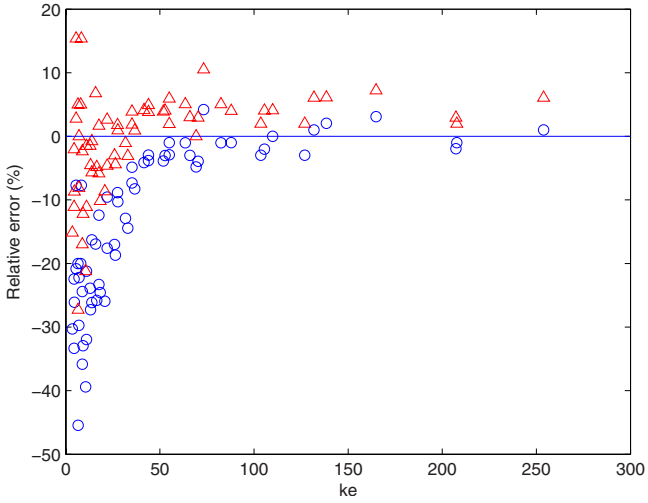


FIG. 12. (Color online) A relative error of the weighted absorption coefficient as a function of ke. (○) Averaged weighting and (△) upper curve.

The relative error of the angle-weighted absorption coefficient is illustrated in Fig. 12. The measured value is used as a reference in defining the relative error. One can see the clear tendency that the error of the averaged weighting function (hollow circle) decreases with increasing ke. The error of upper curve is smaller for low ke, but the error converges to 7% biased from the measurement, as ke becomes higher. One can also observe that the error of the averaged weighting seldom is positive (except five cases), which means the measured value is generally higher than the angle-weighted absorption by the averaged weighting. If ke is higher than 50, the results by the averaged weighting will be acceptable. The upper curve shows smaller errors for low ke, but the error is still too large to be acceptable (the maximum error is 23%). Larger errors at low ke may originate from neglecting diffraction phenomenon in the calculation. It should be mentioned here that the distribution of measured intensity at lower frequency in a reverberant condition is flatter than the intensity distribution at high frequencies in Ref. 15. This fact also supports the better agreement employing the more uniform energy distribution at low frequencies.

It can be summarized that non-uniform energy density can reduce discrepancies between the measured absorption coefficient and the theoretical value when it is taken into account with the size correction. A high value of ke guarantees an acceptable precision by using the averaged weighting, which is similar to a Gaussian distribution.

VI. CONCLUSIONS

The angular distribution of energy density incident on a sample has been simulated for a rectangular room and a reverberation chamber with non-parallel surfaces by using the beam tracing method. A large variation in incident energy density was found depending on the source position. To achieve a uniform distribution, the source should be located perpendicularly from the boundary of the target surface, as close as possible to the target surface. Therefore a room with non-parallel walls is advantageous for obtaining a uniform distribution. A long distance from a source to a target surface

results in a concentration of acoustic energy near the normal direction. The simulated reverberant energy distribution plays the role of a weighting factor in calculating the angle-weighted absorption coefficient. The importance of non-uniform incident energy becomes significant for high ke values. For smaller values of ke, the calculated absorption coefficient adopting fairly uniform distribution agrees well with the measurement, while the averaged Gaussian-like weighting function agrees better with the measurement for high ke.

ACKNOWLEDGMENTS

The author gratefully acknowledges the valuable comments and encouragement from Professor Finn Jacobsen at DTU. The author also would like to thank Dr. Jens Holger Rindel, Dr. Sven-Ingvar Thomasson, Dr. Hyun-Ju Kang, and Professor Jeong-Guon Ih for constructive discussion.

[1]T. J. Schultz, "Diffusion in reverberation rooms," J. Sound Vib. **16**, 17–28 (1971).
[2]R. V. Waterhouse, "Interference patterns in reverberant sound fields," J. Acoust. Soc. Am. **27**, 247–258 (1955).
[3]L. L. Beranek, *Noise Reduction* (McGraw-Hill, New York, 1971).
[4]R. E. Jones, "Inter-comparisons of laboratory determinations of airborne sound transmission loss," J. Acoust. Soc. Am. **66**, 148–164 (1979).
[5]E. T. Paris, "On the reflection of sound from a porous surface," Proc. R. Soc. London, Ser. A **115**, 407–419 (1927).
[6]D. Olynyk and T. D. Northwood, "Comparison of reverberation room and impedance tube absorption measurement," J. Acoust. Soc. Am. **36**, 2171–2174 (1964).
[7]A. de Bruijn, "Influence of diffusivity on the transmission loss of a single-leaf wall," J. Acoust. Soc. Am. **47**, 667–675 (1970).
[8]T. T. Wolde, "Measurements on the edge-effect in reverberation rooms," Acustica **18**, 207–212 (1967).
[9]W. Kuhl, "Der einfluss der kanten auf die schall-absorption poroser materialien (The influence of the edges on sound absorption of the porous materials)," Acustica **10**, 264 (1960).
[10]T. W. Bartel, "Effect of absorber geometry on apparent absorption coefficients as measured in a reverberation chamber," J. Acoust. Soc. Am. **69**, 1065–1074 (1981).
[11]C. W. Kosten, "International comparison measurement in the reverberation room," Acustica **10**, 400–411 (1960).
[12]P. E. Sabine, "Specific normal impedances and sound absorption coefficients of material," J. Acoust. Soc. Am. **12**, 317–322 (1941).
[13]A. de Bruijn, "Influence of diffusivity on the transmission loss of a single leaf wall," J. Acoust. Soc. Am. **47**, 667–675 (1970).
[14]H.-J. Kang, J.-G. Ih, J.-S. Kim, and H.-S. Kim, "Prediction of sound transmission loss through multilayered panels by using Gaussian distribution of directional incident energy," J. Acoust. Soc. Am. **107**, 1413–1420 (2000).
[15]H.-J. Kang, J.-G. Ih, J.-S. Kim, and H.-S. Kim, "An experimental investigation on the directional distribution of incident energy for the prediction of sound transmission loss," Appl. Acoust. **63**, 283–294 (2002).
[16]Y. Makita and K. Fujiwara, "Effects of precision of a reverberant absorption coefficient of a plane absorber due to anisotropy of sound energy flow in a reverberation room," Acustica **39**, 331–335 (1978).
[17]Y. Makita and T. Hidaka, "Revision of the cos theta law of oblique incident sound energy and modification of the fundamental formulations in geometrical acoustics in accordance with the revised law," Acustica **63**, 163–173 (1987).
[18]S.-I. Thomasson, "On the absorption coefficient," Acustica **44**, 265–273 (1980).
[19]S.-I. Thomasson, "Theory and experiments on the sound absorption as function of the area," Report No. TRITA-TAK 8201, KTH, Stockholm, Sweden, 1982.
[20]A. London, "The determination of reverberant sound absorption coefficients from acoustic impedance measurements," J. Acoust. Soc. Am. **22**, 263–269 (1950).
[21]B. S. Atal, "A semi-empirical method of calculating reverberation chamber coefficients from acoustic impedance value," Acustica **9**, 27–30 (1959).
[22]P. Dämmig, "Model investigation into sound fields in reverberation rooms," Acustica **75**, 105–120 (1991).
[23]T. Lewers, "A combined beam tracing and radiant exchange computer-model of room acoustics," Appl. Acoust. **38**, 161–178 (1993).
[24]A. Wareing and M. Hodgson, "Beam-tracing model for predicting sound field in rooms with multilayer bounding surfaces," J. Acoust. Soc. Am. **118**, 2321–2331 (2005).
[25]C.-H. Jeong, J.-G. Ih, and J. H. Rindel, "An approximate treatment of reflection coefficient in the phased beam tracing method for the simulation of enclosed sound fields at medium frequencies," Appl. Acoust. **69**, 601–613 (2008).
[26]I. A. Drumm and Y. W. Lam, "The adaptive beam-tracing algorithm," J. Acoust. Soc. Am. **107**, 1405–1412 (2000).
[27]T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, and A. Ngan, "A beam tracing method for interactive architectural acoustics," J. Acoust. Soc. Am. **115**, 739–756 (2004).
[28]ISO, "Acoustics-measurement of sound absorption in a reverberation room," ISO 354 (International organization for Standardization, Geneva, 1985).
[29]"ASME C 423–02: Standard test method for sound absorption and sound absorption coefficients by the reverberation room" (American Society of Mechanical Engineers, New York, 2003).
[30]J. H. Rindel, "Modeling the angle-dependent pressure reflection factor," Appl. Acoust. **38**, 223–234 (1993).
[31]C. W. Kosten, "Die messung der schallabsorption von materialien" (The measurement of sound absorption of the materials)," in Proceedings of the Third International Congress on Acoustics, Stuttgart, Germany (1959), pp. 815–830.

Increase in voice level and speaker comfort in lecture rooms

Jonas Brunskog[a] and Anders Christian Gade
Acoustic Technology, DTU Electro, Technical University of Denmark, Kongens Lyngby DK-2800, Denmark

Gaspar Payá Bellester
C/Sènia 1, ler. Monòver, E-03640 Alicante, Spain

Lilian Reig Calbo
C/Aitana, 48, BJ-A Muro de Alcoy, E-03830 Alicante, Spain

(Received 11 March 2008; revised 20 January 2009; accepted 20 January 2009)

Teachers often suffer from health problems related to their voice. These problems are related to their working environment, including the acoustics of the lecture rooms. However, there is a lack of studies linking the room acoustic parameters to the voice produced by the speaker. In this pilot study, the main goals are to investigate whether objectively measurable parameters of the rooms can be related to an increase in the voice sound power produced by speakers and to the speakers' subjective judgments about the rooms. In six different rooms with different sizes, reverberation times, and other physical attributes, the sound power level produced by six speakers was measured. Objective room acoustic parameters were measured in the same rooms, including reverberation time and room gain, and questionnaires were handed out to people who had experience talking in the rooms. It is found that in different rooms significant changes in the sound power produced by the speaker can be found. It is also found that these changes mainly have to do with the size of the room and to the gain produced by the room. To describe this quality, a new room acoustic quantity called "room gain" is proposed. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3081396]

PACS number(s): 43.55.Hy, 43.70.Aj [NX] Pages: 2072–2082

I. INTRODUCTION

The primary means of communication in most educational settings are speech and listening. The acoustics of the lecture room can restrict or support the speaker and improve the sound of the voice and the intelligibility of speech. The room acoustics in lecture rooms is therefore an important issue when considering the productivity and working environment in schools and other teaching situations. Thus, a large amount of work has been carried out within this field. However, the large body of published articles focuses on the point of view of the listener. It is therefore easy to find works on speech intelligibility in the room and advisable reverberation times (RTs) and background noise levels (BNLs) in order to achieve good learning conditions, etc., (see, e.g., Bistafa and Bradley[1]). There are also standards and recomendations,[2–4] indicating how well established this field is.

However, it is known that teachers often suffer from health problems or tension related to their voice. Recent works made it evident that the teacher's labor is one of the professions with high vocal demands.[5] Examples of other professions with high vocal demands are actors, singers, journalists, telephone operators, and military personnel. Studies show that a majority of teachers have experienced vocal problems, about one-tenth have severe problems, and 5% have experienced such severe, numerous, and frequent voice problems that their working ability is challenged.[5] For the teacher, in the long run, this voice load due to speaking in the classroom can result in voice disorders such as hoarseness and voice fatigue and can even force teachers to retire early from their profession. Lubman and Sutherland[6] disclosed that this is an important economic problem for governments and private schools.

Most teachers have probably experienced that different rooms vary in comfort when one speaks in them. However, even though the vocal problem is so important, just a few studies about the speaker and his behavior in and impression of the lecture room have been accomplished. One example is Kleiner and Berntson,[7] where the early reflections of the sound produced by the speaker were studied in a synthetic experimental setup. A system of loudspeakers in an anechoic chamber was used to simulate different rooms. All settings simulated rooms with different shapes but the same volume. The interest was in the effect of lateral and vertical early reflections on the speakers' comfort. Different combinations of delayed simulated reflections were tested. A paired-comparison test was used in order to find the setting preferred by the speakers. It was concluded that symmetrical settings were preferred over asymmetrical ones. There was however no significant difference between the different symmetrical settings, and perfectly symmetrical settings are not realistic in real rooms with a movable speaker. It can be noted that this was an entirely subjective study—no objective values were calculated from the simulated impulse responses. Kob *et al.*[8] presented results from a study where the voice status of 25 teachers were investigated using standard methods as applied by audimetrisists, phoniatricians, and speech therapists, in addition to an acoustic analysis of speech and voice samples. The acoustics of some rooms was

[a]Electronic mail: jbr@elektro.dtu.dk

 0001-4966/2009/125(4)/2072/11/$25.00

also investigated, and the result of speaking in different rooms was analyzed dependent on the voice status. The results indicate an influence of both the room acoustics and the voice status on the voice quality of the teachers. But the study used RT and speech transmission index as the parameters describing the room acoustic environment. Thus, no clear distinction was made between the problem perceived by the listener and the speaker.

Several studies in which different voice parameters were measured in real classrooms have been reported, e.g., Rantala *et al.*[9,10] or Jonsdottir *et al.*[11] However, in these studies the influence of the room was not included. Instead, the focus here was to study different subgroups of speakers, e.g., with and without voice problems. The voice parameters were primarly the voice level [defined as the sound pressure level a distance of 1 m from the speaker] and pitch (more specifically the fundamental frequency F_0 of the voice signal) and fluctuations in these parameters.

Thus, the literature relating the room to the speaker and the voice signal produced is rather thin; not much information is available on how to design or improve the room in order to make a better environment for the speaker. However, such information is available in the field of acoustics of rooms for music performance. Also here, the majority of works deal with the conditions for the audience, but there have also been studies concerning how musicians experience and react to the room acoustics. An important example is Gade,[12] who, in a laboratory experiment in an anechoic chamber equipped with a loudspeaker system similar to that of Kleiner and Berntson,[7] let musicians play in and react to simulated sound fields. Gade[13] also carried out corresponding subjective and objective studies in real concert halls. In both cases the subjective response answered by the musicians were correlated with different objective measures. Gade found that the "support" provided by the room—the sensation that the room responds to his instrumental effort—is important for the musicians. Gade defines an objective measure, called *ST*, that correlates well with the sensation of support. *ST* is determined as

$$ST = 10 \log \frac{E_{20-x}}{E_{\text{dir}}}, \tag{1}$$

where E_{20-x} is the energy in the impulse response from 20 ms to x ms (x being either 100 or 200 ms, or even infinity) [see Eq. (2)] and E_{dir} is the energy in the direct path, defined as $E_{\text{dir}}=E_{0-10}$, which is the energy within the first 10 ms. The impulse response is to be measured with a source-receiver distance of 1 m. Obviously, 1 m distance is larger than the typical distance between the musician's ear and his instrument, but this distance was still chosen to obtain a measure with sensible variation and dynamic range. *ST* is thus the fraction of energy coming later than 20 ms relative to the direct sound. In the absence of reflected sound *ST* equals $-\infty$ dB, and a zero support, $ST=0$ dB, means that the total contribution from the reflections equals the direct sound. This definition works well in large rooms where the direct part of the impulse response is clearly separated from the reflections, but measurements of ST is problematic for smaller rooms. Another problem with the definition in Eq. (1) is that it does not clearly reflect what happens close to the source, which at the same time is the position to be studied. In the real situation, e.g., in case of singing or speaking, the source is the mouth and the receiver position is the ear, just a few centimeters away. The direct path is thus described by the transfer function (or impulse response) from the mouth, around the head, to the ear in absence of reflections. How to deal with this is not obvious in case of the definition in Eq. (1). A third problem is that an anechoic chamber is included in the present study, and *ST* is undefined in such a room. Thus, in the present study we have made use of another definition using the measured impulse response of a setup with an artificial dummy head torso and taking as reference the measured value in an anechoic room. The new quantity is called room gain, with abbreviation RG and variable G_{RG} (see Sec. II C).

It seems likely that the vocal problems of teachers are due to the voice level being increased in different situations when teachers feel uncomfortable with the environment. The environment here not only includes the physical environment of the lecture room, but also the students and the overall working conditions. There are two hypotheses here, one being that vocal health problems are related to an environment where the speaker feels that he must increase his voice, the other being that the physical environment itself can cause the speaker to increase his voice. Only the latter will be tested in the present paper. The aim of this project was thus to find some of the parameters that cause the speaker to force their voice and situations when it is uncomfortable to speak.

Aspects not taken into account in this study are the influence of the audience, including the background noise (BN) produced by them, the change in voice during the day, the influence of voice problems of the subjects and other aspects related to the subjects (e.g., mood or attitude toward teaching), and the speech intelligibility in the rooms, subjectively or objectively. Moreover, the study only deals with nonamplified voices.

One question is then which objectively measurable parameters to include in the study. Real rooms are to be used, and the focus is on the speaker, not the listener. Thus, the parameters should be related to what the speaker experiences at the position where he speaks. Parameters related to speech recognition and intelligibility are therefore left out. The impulse response contains all information of the transfer path from source to receiver, and most measures can be calculated from it. It is however important that the source and receiver positions are as correct as possible. Parameters that are extracted from the impulse response are the RT and the RG. Parameters not included in the impulse response are those not directly related to the acoustic transfer path—that is, BN and the size of the room. Thus, four basic parameters are chosen to characterize each room—RT, RG, BNL and *volume*. However, different variants of these parameters were tested as well.

In the subjective study, most of the questions were related to the objective parameters. Thus, the subjects were asked about the impression of reverberation and support, as well as background level in the rooms studied. They were also asked about the general impression of speaking in the

TABLE I. The rooms used in the experiments and their objective values. All rooms are located at DTU, Lyngby, Denmark. ◇, number of questionnaire answers for each room.

Name	Abbrev.	V (m^3)	T_{30} (s)	T_{EDT} (s)	G_{RG} (dB)	$L_{BN,A}$ (dB)	nr.◇
Auditorium 81	A81	1900	1.06	1.12	0.28	41.8	14
Auditorium 21	A21	1220	1.53	1.72	0.29	53.5	19
Lecture r. 019	LR	190	0.46	0.40	0.42	47.5	21
Meeting r. 112	MR	94	0.42	0.33	0.58	47.5	17
Large anechoic ch.	ACH	1000	0.06	0.01	0	45.9	17
IEC listening room chamber	IEC	100	0.34	0.32	1.12	46.7	16

room and if they raised their voice when speaking. A question about echo phenomena was also included in order to be able to say if this parameter influences the general impression of the room.

The main findings in this paper is that the different rooms significantly change the sound power produced by the speaker. It is found that these changes mainly have to do with the size and the RG of the room.

II. METHOD

A. Method overview

Both subjective responses and objective measures of the room and of the voice level are collected. A selection of different natural acoustic environments are used—opposite of using a synthetic sound field. In simulated sound fields the variables can be changed rapidly and with precision in wide ranges. However, the sound quality is still limited due to the need for real time processing of the signals produced by the speaker. Moreover, the visual impression of the room cannot easily be included—this might be a positive aspect in many cases, but here it is important to get the visual size of the room and the distance to the audience right. Therefore, real rooms were chosen to be used—six in total. The range in the physical parameters of the rooms used was wide, including small meeting and listening rooms, a medium size lecture room, two larger auditoria, one with high RT and one with low RT, and a large anechoic room.

In the six rooms the sound power level produced by six speakers was measured. Each of the speakers held a short lecture (about 5 min). Objective room acoustic parameters were measured in the rooms as well, and a subjective questionnaire was handed out to about 20 persons who had experience in speaking in the rooms. A statistical analysis was then used to find relationships between the subjective responses and the objective measures.

B. The subjects

In the objective study six speakers were used. Three of these were teachers at Acoustic Technology, Ørsted*DTU; the other three were students in acoustics. In one of the rooms (meeting room 112, building 352), only five speakers were present. The speakers had no known voice pathology. Each speaker was instructed to give the same lecture in all rooms. However, as the speakers did not have a written text to read, the lectures were not identical. Most speakers used a laptop computer with a Powerpoint presentation as the basis of the speech. In order to get the background level identical, a laptop and a video projector (if available in the room) were present also for those not using it. All speakers were male, age about 20–55. There is a possibility that the speakers do not fully represent all relevant speakers, as it consisted of those finding it interesting to participate. Actually, the teachers participating were known to have weak voices (low voice power). However, most of the analysis are made on a relative voice power level (VPL) (see Sec. II C), which decreases the variance in the data. Another subset problem might be that all subjects were acousticians, a fact that might influence the result—we choose to believe that this has a minor influence only.

In the subjective study 21 subjects participated (between 14 and 21 responses were collected for each room, see Table I). The subjects were teachers and students in acoustics—the participants in the objective part were also present in the subjective part. Both male and female subjects aged between about 20 and 60 participated.

C. Objective measurements and equipment

1. Impulse response measurements

The impulse response $h(t)$ of the rooms is measured to calculate RT and RG. The equipment used for the measurements were power amplifier LAB 300 from LAB Gruppen, microphone unit type 4192-L-001 Brüel & Kjær (B&K), conditioning preamplifier Nexus type 2690 B&K, and sound level calibrator type 4231 B&K. In case of the reverberation measurements, an omnidirectional dodecahedron loudspeaker was used, and in case of the RG measurements a dummy head torso was used, as described below. The DIRAC software[14] was used with e-sweep excitation signal. The sweep length was 21.8 s.

2. Reverberation time

Generally, the most important room acoustic parameter is the RT (variable T_{30}) (see ISO 3382).[15] The *early decay time* (EDT) (variable T_{EDT}), is the RT determined from the first 10 dB range of the decay curve. The EDT is known to be more closely related to the subjective impression of reverberation than RT. In the analysis EDT was mainly used. (A reference of these basic room acoustic parameters is Kuttruff.[16])

The RT is calculated from the impulse response using the Schroeder method.[16] The RTs were calculated in octave

bands. In order to describe the RT as a single number, the arithmetic mean of the RT in the octave bands centered in 500 and 1000 Hz is used.

3. Room gain

The transmission path from the mouth to the ear has three parts: bone conduction, a direct airborne part, and a room reflection part; it is the latter path that is of interest here. The perceived beneficial increase in the loudness caused by the room is assumed to be due to the early reflections as compared to the direct response without reflections, perceived as one's ability to hear oneself properly in the room. This is here denoted as a gain, or support, caused by the room. The parameter used in the present study is called RG (variable G_{RG}). It is defined as the energy (in decibels) of the impulse response measured between the mouth and the ear of a dummy head torso, taking as reference the corresponding measurement in the anechoic chamber where only the direct sound is present. As explained earlier, the reason for not using the support measure *ST* is that small rooms are also to be included in the present study, and then the definition of the *ST* is not appropriate, as the direct part of the impulse cannot be separated from the rest of the impulse response. Moreover, an anechoic chamber is included in the study, and here $ST=-\infty$.

The energy of an impulse response in a time interval t_1 to t_2 can be calculated as

$$E_{t_1-t_2}=\int_{t_1}^{t_2} h^2(t)dt, \tag{2}$$

where $h(t)$ is the impulse response. The energy in the entire impulse response is in the same way,

$$E=\int_0^{\infty} h^2(t)dt. \tag{3}$$

The corresponding impulse energy level is $L_E=10\log E/E_{\mathrm{ref}}$, where E_{ref} is the reference value. The RG is then defined as the energy in decibel in the signal relative to the direct energy as measured in the anechoic chamber,

$$G_{\mathrm{RG}}=L_E-L_{E,\mathrm{ach}}=10\log E/E_{\mathrm{ach}}, \tag{4}$$

where $L_{E,\mathrm{ach}}$ and E_{ach} are the impulse energy level and energy in the anechoic chamber, respectively.

The RG is related to the support *ST*, as defined in Eq. (1). If it is assumed that $E_{\mathrm{dir}}\approx E_{0-20}\approx E_{\mathrm{ach}}$ and $E_{20-x}\approx E_{20-\infty}$, then

$$ST\approx 10\log\frac{E-E_{0-20}}{E_{\mathrm{ach}}}\approx 10\log(10^{G_{\mathrm{RG}}/10}-1). \tag{5}$$

A support value of $ST=0$ thus corresponds to $G_{\mathrm{RG}}=3$ dB, meaning that the reflections contribute with the same energy as the direct sound. It should, however, be noticed that the source/receiver distance is different in the definition of *ST* as compared to G_{RG}.

The equipment used was the same as described under the impulse response above, with the following changes: dummy head, head and torso simulator type 4128 with right ear simulator type 4158 and left ear simulator type 4159 B&K, and power amplifier for the sound source (the dummy mouth).

The dummy head was placed in the area where the speaker normally stands during the lecture (next to the blackboard or similar). The average of six different positions of the dummy head was used. Moreover, the average RG of the left and right channels was calculated and used in the data analysis.

The RG was calculated from the impulse response by means of postprocessing in MATLAB. All signals have been normalized with a maximum amplitude of the signal to 1 (amplitude of the first peak of the impulse response). Some problems with the signals were found during the analysis. Noise was found in all the signals. In order to reduce the effect of this problem, all the impulse response signals were truncated (cutted) so as to avoid the last part of the signal, which mainly contained noise. Thus, the noise effect was minimized, and it is judged that its influence can be disregarded.

The RG was calculated per octave band. In order to define the RG of the room with one characteristic value, the arithmetic mean of the RG in the octave bands between 125 Hz and 4 kHz is used.

4. Background noise level

In a speech situation the BNL (variable $L_{BN,A}$) is important. BNL can be defined as the sound pressure level of the noise measured in the absence of the sound under investigation—in this case the speech. The BN can originate from the ventilation systems, the outdoor environment and traffic, equipment such as computers and projectors, and the students/audience. As the BNL increases, the speaker may increase his voice to compensate and overcome the noise in order to be heard. The voice will be affected by the mental pressure due to the failure of being heard. The frequency content in the voice signal will then be changed—there will be more high frequency content due to an increased fundamental frequency. These changes are known as the Lombard effect; an early reference is Lane and Tranel.[17] The effect is included in ANSI-S3.5.[2] (Sometimes, the term "Lombard effect" is restricted to just the increase.) This is also closely related to the fact that in a situation with several people talking to each other, they increase their voice to overcome the BNL that is produced by all the persons speaking, producing a nonlinear feedback loop, see, e.g., Hodgson *et al.*[18] Naturally, the number of students and their behavior during the lecture also may play an important role here—the students will contribute to the background level and probably react in relation to the Lombard effect. However, this aspect is not part of the present work (due to schedule reasons and time limits); the present project is focused on the characteristics of the room only, leaving this important aspect to further research. The number of listeners present in the room was just a few (3–5) and adult, so there contribution to the BNL is assumed to be low. The BNL naturally present in the rooms (from the ventilation system, video projector, computers, etc.) was, however, registered.

The equipment used to measure L_{BN} is the same as for the impulse response measurements for the reverbation. The measurement duration is 21.8 s. The mean value of six microphone positions have been used in all rooms. The positions were in the area the teacher was using. To get a single value, the A-weighted level $L_{BN,A}$ is used. The equipment used by the speakers (laptop computer and projector) was present in the room during the measurement.

5. Room volume

Of the objective parameters describing the rooms, finally the size or *volume* (variable V) has also been used. The hypothesis here is that the speakers unconsciously adjusts the level of the voice depending on the room size and the distance to the audience, so that everyone is likely to hear. However, it is not clear if it is the volume by itself or a typical length scale in the room that is the primary variable here. Thus, V, $\log V$, and $\sqrt[3]{V}$ were all tested.

6. Voice power level

With the rooms defined, the last step is to define the behavior of the speaker in the room. In this project, this is described by the strength of the speaker's voice. The quantity used here is the *voice power level* (VPL) (variable L_W) that is the source power in decibel. Thus, the sound power level produced during speech by the different test speakers was measured in the different rooms.

The measurement of the VPL is a central issue of this paper. The measurements are made with a computer phone conversation headset, placed on the speaking subjects. The experimenter made sure that the position of the headset was fixed to the same position in all measurements, about 3 cm from the mouth. The equipment consisted of Headset Creative HS-390 and sound analyzer DIRAC. The signals were measured while the speaker was lecturing for about 5 min. An average of 15 signal segments of 21.8 s were used for each subject.

A calibration procedure was needed to transfer the measured signals to sound power level L_W. The dummy head torso equipped with a loudspeaker in the mouth was placed in a reverberation chamber with the headset attached in the same position as described above. A broad band noise signal was fed to the loudspeaker and measured simultaneously by the headset and with microphones in the reverberant field of the room according to sound power level standard measurements (ISO 3743-2). The measurements and calibrations were performed in octave bands. The relation between the sound power of a source and the sound pressure level in one position determined by a microphone can generally be expressed as $L_W=L_p+G$, where G is a gain constant for the setup (depending on the source-receiver distance and source directivity) and L_p is the sound pressure level as measured by the headset. It is now assumed that the microphone is so close to the source that only the direct field is present (i.e., the signal to noise ratio is assumed to be so good that the room response can be neglected). Moreover, it is also assumed that all speakers had the same directivity, equal to that of the dummy head. It is thus assumed that G is constant during all measurements in all rooms. (Note that this quantity obviously is different from G_{RG}.) Finally, having determined both L_W and L_p at the same time in the reverberation chamber, the gain constant G is determined.

The VPL is determined in octave bands from 125 Hz to 4 kHz. In order to have a single value, three different methods are tested: linear ($L_{W,l}$) and A-weighted ($L_{W,A}$) absolute VPL and linear VPL relative to the VPL in the anechoic chamber (ACH), ΔL_W. Note that the subtraction is made for each speaker, so that ΔL_W is made relative to the VPL for that speaker in the ACH. In this way the variance is reduced. The ACH room was chosen as it was the room with the highest average VPL. (The room with the lowest VPL, the meeting room (MR), was also considered to be used as a reference, but this idea was dropped as not all speakers spoke in this room.)

D. The rooms

To get good statistic results, it is important to apply a wide range and even distribution of the different physical variables defining the room. The rooms and the values of the objective measures are given in Table I. The rooms were a small MR and an IEC listening room (IEC), a medium size lecture room (LR), two larger auditoria, one with high RT (A21) and one with low (A81), and a large anechoic room (ACH). Including the anechoic room means that the subjects have a very clear reference for RT and RG—which both are zero in this room. Besides, ACH is relevant as it represents out door surroundings. The range covered by the volume, the RT and the RG can be considered large in comparison to what can be found in real life situations. For the BN, only the naturally present BN was included. Thus, this variation is small as compared to what can be found in real life situations.

E. Questionnaire and subjective response

In an attempt to relate the objective parameters of the room and the VPL to the subjective experience of the rooms, a questionnaire was designed. The questions were formulated after a first interview with a few teachers. The parameters considered are described below.

The questions were answered for each of the rooms in which the subject had experience talking. Thus, the subjects were not necessarily in the room when the questions were answered—in an attempt to increase the number of answered questionnaires. 21 subjects answered the questions; the number of answers for each room varied between 14 and 21 (see Table I). The questions were answered on a scale from 1 to 7. Only the natural numbers were used. Taking the arithmetic average of these answers, a subjective response variable S_i was formed, where the index i is the abbreviation of the question (see below).

The questions are the following (the questions are given in italic fonts)—it should, however, be noted that the these are not exactly the questions used (due to poor English).

Do you consider this room to be good to speak in? This question is referred to the degree of comfort and how easy it is to speak in the room. The rank is between low if the room

TABLE II. Significance test of the subjective response parameters and VPL parameters (different versions) using ANOVA. The following symbols are used: * means significant at the 5% level, *** means significant at the 0.1% level, and — means no significance at the standard levels.

Question	GSI	TR	ECHO	BN	IV	ES	$L_{W,I}$	$L_{W,A}$	ΔL_W
p-value	$<10^{-6}$	$<10^{-6}$	0.046	0.16	$<10^{-6}$	$<10^{-6}$	0.13	0.11	0.036
Significance	***	***	*	—	***	***	—	—	*

is not good to speak in and high if it is good to speak in. This parameter is labeled GSI, variable S_{GSI}.

Do you think the RT is too long in the room? This question clearly refers to the objective parameter of RT. The rank in this case goes from "no" if the reverberation is not too long or "yes" if it is too long. This parameter is labeled TR, variable S_{TR}.

Have you noticed echo phenomena in the room? The sensation of echo might influence the general impression of the room, so this response is introduced even though it is not represented in the objective parameters. The answers should be covered between low if no echo is noticed and high if there is too much echo. This parameter is labeled ECHO, with variable S_{ECHO}. A low score is considered good.

Is the BN too high in the room? The subjects' response might be from "yes" if they think there is a lot of BN in the studied room to "no" if they think that there is no noise in the room. This parameter is labeled BN, variable S_{BN}. A low score is considered good.

Do you have to increase your voice in this room to be heard? This question is interrelated to the sound power level. The answer is between "no" if the subjects think they did not increase the voice, to "yes" if they did have to increase the voice. This parameter is labeled IV, variable S_{IV}. A low score is considered good.

Is there enough support in this room? This has to do with whether the room helps the speaker to hear himself. The rank is between bad support if they believe that the room does not yield support at all and good support if the support is sufficient. This parameter is labeled ES, variable S_{ES}. A high score is considered good.

F. Data analysis

The statistical analysis of the data was carried out in MATLAB. This analysis incorporates analysis of variance (ANOVA), correlation coefficients, and linear regressions.

In order to find relationships between the subjective responses and the objective parameters—a psychometric function—some postprocessing has been done. The psychometric function, relating a subjective parameter S with upper limit $S_{\max}$ and lower limit $S_{\min}$, and an objective parameter d (or a linear combination between such parameters) should be an S-shaped function. The reason for this is that the objective parameter is not bounded, $d \in [-\infty\infty]$, but the subjective parameter is bounded, $S \in [S_{\min}S_{\max}]$. One choice of such a function is

$$S = \frac{S_{\max} - S_{\min}}{1 + e^{-d}} + S_{\min} \tag{6}$$

(this choice of psychometric function is taken from paired-comparison theory[19,20]). The point of using such a relation is that S has a finite domain $S \in [S_{\min} \cdots S_{\max}]$, whereas d might have an infinite domain $d \in [-\infty \cdots \infty]$. In the present case $S_{\max}=7$ and $S_{\min}=1$. Solving for d in Eq. (6), a suitable transformation from the finite S-domain to the infinite d-domain of the objective measures is found,

$$d_S \equiv -\ln \frac{S_{\max} - S}{S - S_{\min}}. \tag{7}$$

The parameter d_S can be used as the dependent variable in regressions connecting objective measures to subjective response.

However, in some cases the objective parameter is nonnegative, $d>0$. That is the case for the RT and the RG. Moreover, in the present study the extreme situation of zero RT and RG is included in the study due to the use of the anechoic chamber. In these cases Eqs. (6) and (7) have to be modified. The following equations then applies:

$$S = \frac{2(S_{\max} - S_{\min})}{1 + e^{-d}} + 2S_{\min} - S_{\max} \tag{8}$$

and

$$d_S \equiv -\ln \frac{S_{\max} - S}{S - 2S_{\min} + S_{\max}}. \tag{9}$$

However, in many cases the range of the objective parameter is so small that the error of using a linear regression directly between d and S is small. That is actually the case in the present study, and in the result section below, the regressions are often performed both using the psychometric function and directly between S and d.

III. RESULTS

A. Validity and quality of the data

An ANOVA is used to examine if the variations in the data are sigificant. The left part of Table II presents these results concerning the subjective parameters. The variations are significant except for BN, where no significant variations are found at the 5% level or better (p-value of 0.16), and for detection of echo ECHO, where the variations are significant at the lower level of 5% (p-value of 0.046), but not higher. It should here be noted that the variation in the background level of the rooms was small and that there are no known problems with echo or flutter echo in the rooms used. In the

TABLE III. The rooms used in the experiments and their objective values.

Abbrev.	$L_{W,l}$ (dB)	$L_{W,A}$ (dB)	ΔL_W (dB)
A81	62.9	60.0	−1.30
A21	63.9	60.9	−0.08
LR	62.9	60.1	−1.93
MR	58.7	55.2	−4.33
ACH	65.0	62.1	0
IEC	59.8	57.0	−4.32

same way, the right part of Table II presents the significance test of different versions of the VPL. Here the significance of the variations in the data is less, probably due to the lower number of subjects participating. However, taking VPL relative to the result in the anechoic chamber, ΔL_W, yields significant variations at the 5% level (p-value of 0.036).

In the further analysis, only $L_{W,l}$ and ΔL_W will be used describing the VPL. $L_{W,A}$ is disregarded as it does not increase the significance much and is not as straightforward as $L_{W,l}$. Moreover, results depending on the subjective responses BN and absolute VPL, $L_{W,l}$, should be considered only as trends.

B. Relationships among objective parameters

The objective parameters used to describe the rooms were presented in Table I. The objective changes in the VPL are presented in Table III. The correlation matrix between these parameters is given in Table IV. It should be noted that the VPL measures correlate well with the volume, especially log V, and the RG G_{RG}. There is no significant correlation between the VPL measures and RT and BN. It should also be noted that the RT measures and the BN measure do not correlate significantly with any other measure.

Note that the correlation between support ST as calculated in Eq. (1) and the other parameters is not included here as the support is undefined in the anechoic chamber due to the lack of reflections (the value would be $-\infty$).

The results of single variable linear regression are found in Table V. Only results with $p<0.1$ are shown. It is shown once again that log V and G_{RG} correlate well with VPL. A multiple linear regression model using these two variables is

TABLE V. Single variable linear regression. Only regressions with $p<0.1$ are shown. Left: between VPL ΔL_W and the objective parameters. Right: between S_{GSI} and the subjective parameters. The variables b_0 and b_1 are the regression constants, the constant term and the linear term, respectively.

Dependent variable	ΔL_W		S_{GSI}	
Independent variables	log V	G_{RG}	S_{IV}	S_{ES}
R^2	0.78	0.74	0.73	0.61
p	0.02	0.03	0.03	0.07
b_1	2.94	−4.40	−0.90	0.72
b_0	−9.64	−0.021	8.30	1.91

$$\Delta L_W = -5.68 + 1.81 \log V - 2.28 G_{\mathrm{RG}}, \tag{10}$$

with $R^2=0.86$ and $p=0.05$. The improvement of using two parameters is described by the fact that R^2 increases from 0.78 to 0.86 and at the same time the model is at the limit of significance. The model is shown in Fig. 1.

C. Relationships among subjective parameters

The subjective response parameters are presented in Table VI. The correlation matrix for these parameters is given in Table VII. Using the objective domain transformation according to Eqs. (7) and (9) yielded similar results.

The results of single variable linear regressions are found in the right part of Table V. Only results with $p<0.1$ are shown. It can be seen that S_{IV} and S_{ES} correlate well with S_{GSI}; these regressions are also shown in Figs. 2 and 3. A multiple linear regression model using these two variables is

$$S_{GSI} = 6.82 - 0.715 S_{IV} - 0.189 S_{ES}, \tag{11}$$

with $R^2=0.74$ and $p=0.13$. Thus, the improvement of the two parameter model was not large, and the model is not significant. This is probably due to a high linear dependency between S_{IV} and S_{ES}.

D. Relationships between subjective and objective parameters

Table VIII shows the correlation between the objective parameters and the subjective responses (the number of objective parameters has been reduced as T_{30} and $\sqrt[3]{V}$ have been

TABLE IV. Correlation matrix for the objective measures, including the VPL. Only correlations with p-values lower than 0.2 are shown. In parentheses: $0.2>p>0.1$; roman upright: $0.1>p>0.05$; italic: $0.05>p>0.01$; boldface: $p<0.01$.

Objec.	$L_{W,l}$	ΔL_W	T_{30}	T_{EDT}	V	log V	$\sqrt[3]{V}$	L_{BN}	G_{RG}
$L_{W,l}$	**1**	**0.97**	—	—	(0.63)	*0.82*	0.76	—	−0.81
ΔL_W	**0.97**	**1**	—	—	(0.72)	*0.88*	*0.84*	—	*−0.86*
T_{30}	—	—	**1**	**1.00**	—	—	—	—	—
T_{EDT}	—	—	**1.00**	**1**	—	—	—	—	—
V	(0.63)	(0.72)	—	—	**1**	**0.96**	**0.98**	—	(−0.63)
log V	*0.82*	*0.88*	—	—	**0.96**	**1**	**1.00**	—	−0.76
$\sqrt[3]{V}$	0.76	*0.84*	—	—	**0.98**	**1.00**	**1**	—	(−0.72)
L_{BN}	—	—	—	—	—	—	—	**1**	—
G_{RG}	−0.81	*−0.86*	—	—	(−0.63)	−0.76	(−0.72)	—	**1**

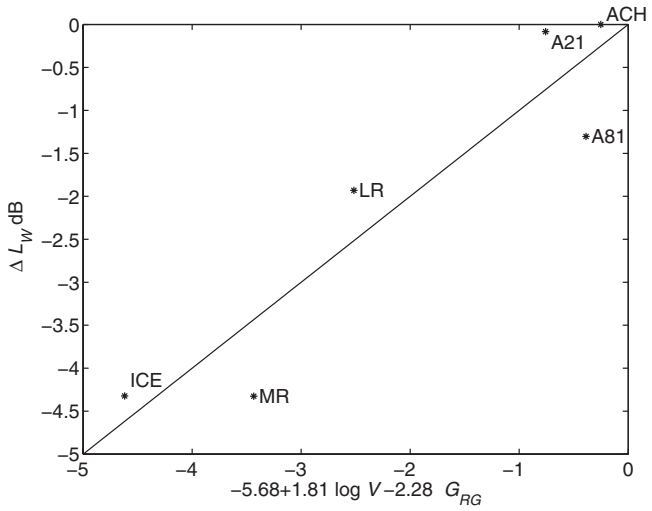


FIG. 1. Regression model (10) versus the real data of increase in VPL ΔL_W. Room abbreviation according to Table I.

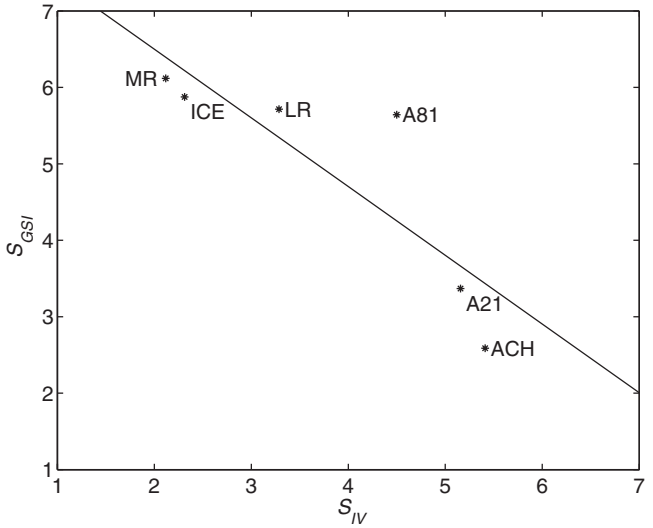


FIG. 2. Regression model between subjective variables S_{GSI} (good to speak in) against S_{IV} (increase voice), according to right part of Table V. Room abbreviation according to Table I.

ignored). Using the objective domain transformation according to Eq. (7) and (9) again yields similar results (a slightly better correlation on average).

The results from single variable linear regressions are found in Table IX. Only the regressions with $p<0.1$ are shown. The regression between IV and ΔL_W is shown in Fig. 4, and that between TR and T_{EDT} is shown in Fig. 5. A multiple linear regression model for IV using two variables is

$$S_{IV} = -0.198 + 1.73 \log V - 1.11 G_{\mathrm{RG}}, \tag{12}$$

with $R^2=0.90$ and $p=0.03$. The improvement of using two parameters is described by the fact that R^2 increases from 0.86 to 0.90 while the model is still significant.

IV. ANALYSIS AND DISCUSSION

The ANOVA test in Table II indicates that in general the statistical quality of the subjective data is better than in the VPL data. One reason for this is probably the higher number of participants in the subjective questionnaire (about 20) as compared to the VPL measurements (about 6). However, it is known that it is difficult to get statistically consistent data for the voice strength (see, e.g., Rantala *et al.*[9]). Anyway, in the present study significant variations in the VPL data are found in case of the relative VPL, ΔL_W, using just six subjects. One reason for this is the normalization procedure of the data by taking the value relative to the anechoic chamber. In this way the natural variation in VPL among the subjects is reduced, and only the increments for different rooms are studied. Moreover, using a wide range of different rooms—including the anechoic chamber, large auditoriums, and small meeting rooms—is likely to increase the variation in VPL.

Considering Table IV, room volume and RG show high correlation with the VPL. An increase in volume increases the VPL, and an increase in RG decreases the VPL. These results are significant if considering ΔL_W related to $\log V$ and G_{RG}. Of the size measures, the logarithm of the volume, $\log V$, has the highest correlation. One can regard $V^{1/3}$ to be a typical length scale of the room and $\log V$ to be related to the average sound pressure level in the room for a given source power level. Thus, the fact that the increase in VPL is better correlated to $\log V$ than $V^{1/3}$ suggests that the aural cues might be more important than the visual cues. The VPL relative to the value in the anechoic chamber, ΔL_W, correlates in general better than the absolute linear VPL, $L_{W,l}$. This is probably linked to the fact that ΔL_W has higher significance than $L_{W,l}$ in the ANOVA test in Table II. Equation (10) expresses the relationship between ΔL_W, $\log V$, and G_{RG}, also shown in Fig. 1. In Table VIII there is a trend that ΔL_W is correlated with ES, the question related to support in the room. Moreover, $\log V$ and G_{RG} are correlated to IV, the

TABLE VI. The rooms used in the experiments and their subjective response values. The scale is between 1 and 7. The notation is $\bar{S}/s$, where $\bar{S}$ is the average value and s is the standard deviation. In the further analysis the average value is used then denoted as S.

Abbrev.	S_{GSI}	S_{TR}	S_{ECHO}	S_{BN}	S_{IV}	S_{ES}
A81	5.64/0.74	2.64/1.34	1.93/1.64	4.00/1.52	4.50/1.34	3.29/0.83
A21	3.37/1.54	5.16/1.50	3.42/2.11	3.74/1.59	5.16/1.26	4.16/0.96
LR	5.71/0.78	1.76/0.54	2.95/2.01	4.33/1.43	3.29/1.27	5.05/0.86
MR	6.12/1.27	2.00/1.00	2.53/2.03	4.59/1.80	2.12/1.05	5.53/0.94
ACH	2.59/2.03	1.00/0	1.41/1.46	5.29/2.73	5.41/2.12	1.29/0.99
IEC	5.88/1.54	1.63/1.08	2.38/2.31	5.06/2.38	2.31/1.01	5.50/0.97

TABLE VII. Correlation matrix for the subjective measures using the subjective scale *S*. Only correlations with *p*-values lower then 0.2 are shown. In parentheses: $0.2>p>0.1$; roman upright: $0.1>p>0.05$; italic: $0.05>p>0.01$; boldface: $p<0.01$.

Subj.	GSI	TR	ECHO	BN	IV	ES
GSI	**1**	—	—	—	*−0.85*	0.78
TR	—	**1**	(0.71)	*−0.84*	—	—
ECHO	—	(0.71)	**1**	(−0.66)	—	0.66
BN	—	*−0.84*	(−0.66)	**1**	—	—
IV	*−0.85*	—	—	—	**1**	*−0.85*
ES	0.78	—	0.66	—	*−0.85*	**1**

question if the subject had to increase the voice to be heard. There is also a trend that log *V* and G_{RG} are correlated to ES. These results confirm the results above.

Considering again Table IV, RT and BNL did not show any correlation with the VPL. Both of these results can seem surprising; RT is the generally most frequently used room acoustic measure, and BN is known to increase the speech level in other circumstances, e.g., in connection with the Lombard effect.[18] However, there is an important difference between these parameters in the present study. The variation in the RT data is rather large, T_{EDT} from 0.01 s in the anechoic room to 1.53 s in auditorium 21, but the variation in background level is small, from 41.8 dB (A) in auditorium 21 to 53.5 dB (A) in auditorium 21 (see Table I). "Large" and "small" should be understood as relative to what is normally found in lecture rooms. Moreover, the BNL in the room used was too low to influence speech. It is thus quite likely that a dependency in BN could be found if more extreme values had been included. The same conclusion does not apply for the RT. Moreover, in Table VIII it can be noted that ΔL_W is not correlated with the corresponding subjective responses TR or BN, which confirms the discussion above.

Considering the correlation among the subjective responses (Table VII), it can be noted that the question of whether the room is good to speak in, GSI, is correlated with the question about increase in voice level to be heard, IV. Thus, the ability to make oneself heard is judged to be important in the general judgment of the room. This is confirmed in Table VIII where GSI is correlated with ΔL_W. There is also a trend that GSI is correlated to ES, the question of whether there is enough support in the room. The other questions (TR, ECHO, and BN) do not show any correlation. It can thus be concluded that a room is good to speak in if it has support, and it is not necessary to increase the voice too much.

In Table VII it can also be seen that the question of whether the RT is too long, TR, is correlated to the question of whether there is too much BN (with negative sign due to the orientations of the subjective scales). Moreover, in Table VIII it is found that also T_{EDT} is correlated to BN but L_{BN} is not. This might seem strange. However, it should be remembered here that the questionnaire was not answered at the same time as the measurements, and that the subjects had the option to answer it while being elsewhere. Thus, BN is rather the experience of the BN as they could remember it. The most severe source of BN is probably the students present during the lecture. In the light of the Lombard effect, it is likely that this noise increases with increasing RT. It is thus not so surprising that T_{EDT} turns out to correlate well with BN. Thus, the subjective response BN does not refer to and is not related to the measured BN.

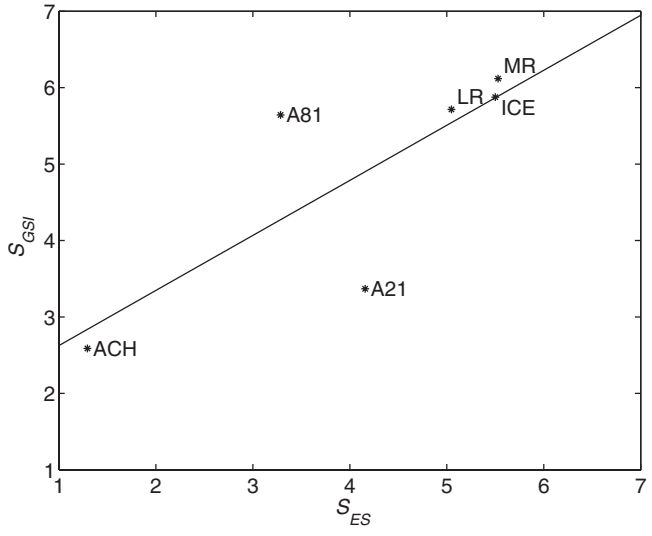


FIG. 3. Regression model between subjective variables S_{GSI} (good to speak in) against S_{ES} (enough support), according to right part of Table V. Room abbreviation according to Table I.

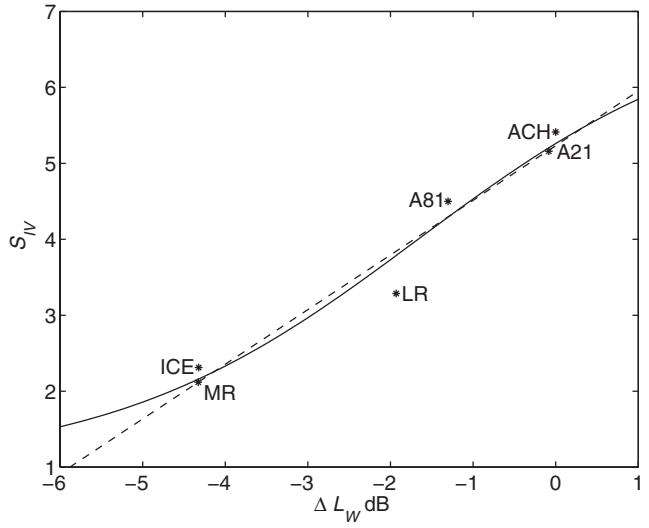


FIG. 4. Regression model between subjective variable S_{IV} (increase voice) against increase in VPL ΔL_W according to Table IX. Room abbreviation according to Table I. Solid line: Using objective domain transformation equation (7). Dashed line: Linear regression.

TABLE VIII. Correlation matrix for the objective and the subjective measures using the subjective scale S. Only correlations with p-values lower than 0.2 are shown. In parentheses: $0.2>p>0.1$; roman upright: $0.1>p>0.05$; italic: $0.05>p>0.01$; boldface: $p<0.01$.

Obj. and subj.	S_{GSI}	S_{TR}	S_{ECHO}	S_{BN}	S_{IV}	S_{ES}
$L_{W,l}$	−0.80	—	—	—	**0.94**	−0.80
ΔL_W	*−0.82*	—	—	—	**0.98**	−0.79
T_{EDT}	—	**0.96**	—	*−0.90*	—	—
V	—	—	—	—	0.79	(−0.65)
$\log V$	(−0.63)	—	—	—	**0.93**	−0.77
L_{BN}	—	(0.65)	0.78	—	—	—
G_{RG}	0.68	—	—	—	*−0.83*	0.80

In Table VII it is also found that there is a trend that the question of whether echo is noticed, ECHO, is correlated to the question of whether there is enough support in the room, ES. This can be interpreted as follows: the reflections that contribute to the RG and support also might be imagined to cause echo phenomena, e.g., flutter echo. However, ECHO does not show big influence on any other parameter and is not correlated with GSI or IV, so it is judged that echo phenomena have not influenced the results. None of the rooms are known to have problems with flutter echo.

In Table VII the question of whether there is enough support in the room, ES, is correlated to the question of whether the subject had to increase the voice to be heard, IV. This seems natural, and it is also reflected in the correlation between ΔL_W and G_{RG} among the objective measurements (Table IV).

The strong correlation between the subjective response of increasing the voice, S_{IV}, and the objectively measured VPL should be noticed in Table VIII. This can be interpreted as the subjects being aware that they have to increase the voice in the room.

In Table VIII T_{EDT} is strongly correlated to TR. Thus, the subjects are aware of the RT. It should then be remembered that all subjects were teachers or students in acoustics and therefore familiar with the concept of RT.

Concerning the frequency rang of RT and RG: the frequency rang used (the octave bands from 125 Hz to 4 kHz for the RG and 500 Hz and 1 kHz octave bands for the RT) has in this study been assumed to be most responsible for the impression of the two measures. Different versions of the parameters have been tested, but not reported, and the chosen definitions and frequency range give good correlation. However, there probably is a need for more research in order to finetune the measures.

Using the regression between ΔL_W and IV (Table IX and Fig. 4), some preliminary design guidelines can be proposed. If a subjective response of $S_{IV} \leqslant 3$ is regarded as a good room, the model yields that this corresponds to $\Delta L_W \leqslant -3.1$ dB. Now, using the model in Eq. (10) (see Fig. 1), this corresponds to $G_{RG} \geqslant 0.80 \log V - 1.1$ dB. Thus, for a room with volume 100 m^3 the RG should be $G_{RG} \geqslant 0.5$ dB, and for a room with volume 1000 m^3 the RG should be $G_{RG} \geqslant 1.3$ dB. It should however be noted that such guidelines are preliminary, and should not be used before further evidence has been obtaind. Also note that the recommended values might be difficult to realize in reality for large auditoriums. Thus, these guidelines are limited to smaller rooms and rooms without voice amplification systems.

V. CONCLUSIONS

The voice power relative to the value in the anechoic chamber varies significantly between room.

The increase in the voice power produced by a speaker lecturing in a room is correlated with the size of the room (especially $\log V$) and the gain produced by the reflections in the room, G_{RG}. These relations are significant.

No significant correlation is found between the increase in the voice power and the RT or background level of the

TABLE IX. Single variable linear regression between subjective and objective variables. Only regressions with $p<0.1$ are shown. The upper part uses the subjective domain S, and the lower part uses the objective domain d_S according to Eqs. (7) and (9).

Dependent variable	S_{GSI}	S_{TR}	S_{IV}		
Independent variables	ΔL_W	T_{EDT}	ΔL_W	$\log V$	G_{RG}
R^2	0.68	0.92	0.96	0.86	0.69
p	0.04	0.003	0.0006	0.007	0.04
b_1	−0.64	2.20	0.72	2.27	−3.13
b_0	3.61	0.94	5.23	−2.12	5.20
R^2	0.71	0.89	0.97	0.86	0.69
p	0.03	0.005	0.0004	0.008	0.04
b_1	−0.50	0.903	0.538	1.68	−2.32
b_0	−0.27	−0.075	0.895	−4.55	0.863

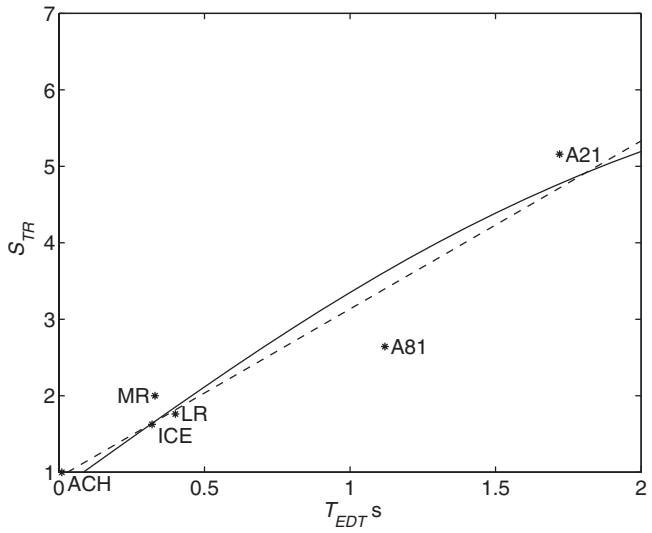


FIG. 5. Regression model between subjective variable S_{TR} (reverberation) against early RT T_{EDT} according to Table IX. Room abbreviation according to Table I. Solid line: Using objective domain transformation equation (9). Dashed line: Linear regression.

room in this study. The latter is probably due to the too small variations in the background levels in the rooms studied.

The general impression of whether a room is good to speak in is linked to the impression of whether it is necessary to increase the voice in the room and if the room provides support to the speaker. The former relation is significant, and the latter is only a trend.

There is a significant correlation between the question of whether the subject had to increase the voice and the actual increase in voice power. There is also a significant correlation between the question about the reverberation in the room and the measured RT. This means that the subjects participating were aware of these parameters.

ACKNOWLEDGMENTS

The authors are grateful to the colleagues and students at DTU participating in the study. Most of the practical work presented in this paper were carried out by two students at DTU during their MSc projects, Gaspar Payá Bellester and Lilian Reig Calbo.

[1]S. R. Bistafa and J. S. Bradley, "Reverbation time and maximum background-noise level for classrooms from a comparative study of speech intelligibilty metrics," J. Acoust. Soc. Am. **107**, 861–875 (2000).
[2]ANSI-S3.5, Methods for the calculation of speech intelligibility index, ANSI, American National Standard Institute, New York, 1997.
[3]*Building Bulletin 93, Acoustic Design of Schools: A Design Guide*, edited by C. Hopkins (Dep. Education and Skills, London, UK, 2004).
[4]IEC-60268-16, Sound system equipment—Part 16: Objective rating of speech intelligibility by speech transmission index, IEC, International Electrotechnical Commission, Geneva, 1998.
[5]E. Vilkman, "Voice problems at work: A challenge for occupational safety and health arrangement," Folia Phoniatr Logop **52**, 120–125 (2000).
[6]D. Lubman and L. C. Sutherland, "Good classroom acoustics is a good investment," in *17th ICA Proceedings* (Universita di Roma "La Sapienza," Rome, Italy, 2001), Vol. **V**, pp. 138–139.
[7]M. Kleiner and A. C. O. Berntson, "Early frontal plane reflections preferred for talkers," in *12th ICA Proceedings* (Canadian Acoustical Association, Toronto, Canada, 1986).
[8]M. Kob, G. Behler, A. Kamprolf, O. Goldschmidt, and C. Neuschaefer-Rube, "Experimental investigations of the influence of room acoustics on the teacher's voice," Acoust. Sci. & Tech. **29**, 86–94 (2008).
[9]L. Rantala, L. Paavola, P. Körkkö, and E. Vilkman, "Working-day effects on the spectral characteristics of teaching voice," Folia Phoniatr Logop **50**, 205–211 (1998).
[10]L. Rantala, E. Vilkman, and R. Bloigu, "Voice changes during work: Subjective complaints and objective measurements for female primary and secondary schoolteachers," J. Voice **16**, 344–355 (2002).
[11]V. Jonsdottir, A.-M. Laukkanen, and E. Vilkman, "Changes in teachers speech during a working day with and without electric sound amplification," Folia Phoniatr Logop **54**, 282–287 (2002).
[12]A. C. Gade, "Investigations of musicians' room acoustic condition in concert halls. Part I: Method and laboratory experiments," Acustica **69**, 193–203 (1989).
[13]A. C. Gade, "Investigations of musicians' room acoustic condition in concert halls. Part II: Field experiments and synthesis of results," Acustica **69**, 249–262 (1989).
[14]Brüel &Kjær, Product data, Dirac Room Acoustics Software, type 7841, www.bksv.com (Last viewed 2/21/09).
[15]ISO-3382, Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters, ISO, International Organization for Standardization, Geneva, 1997.
[16]H. Kuttruff, *Room Acoustics*, 4th ed. (Spoon, London, 2000).
[17]H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709 (1971).
[18]M. Hodgson, G. Steininger, and Z. Razavi, "Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments," J. Acoust. Soc. Am. **121**, 2023–2033 (2007).
[19]P. V. Rao and L. L. Kupper, "Ties in paired-comparison experiments: A generalization of the Bradley-Terry model," J. Am. Stat. Assoc. **62**, 194–204 1967.
[20]A.-C. Johansson, P. Hammer, and E. Nilsson, "Prediction of subjective response from objective measurements applied to walking sound," Acta. Acust. Acust. **90**, 161–170 (2004).

A general wave decomposition formula for the measurement of normal incidence sound transmission loss in impedance tube

Yacoubou Salissou[a] and Raymond Panneton
GAUS, Department of Mechanical Engineering, Université de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada

(Received 14 October 2008; revised 16 January 2009; accepted 22 January 2009)

Two types of general methods can be found in the literature for the determination of the normal incidence sound transmission loss (nSTL) of acoustical elements. The first one is based on the transfer matrix (TM) approach, and the second one is based on the wavefield decomposition (WD) theory. From all the techniques proposed in the literature, the general TM methods (two-load or two-source location) are the only methods yielding the exact nSTL of an acoustical element without any assumptions on its symmetry and on the termination (i.e., the load). Except for the case of an anechoic termination, there is no method based on the WD theory which yields exact nSTL. This paper presents a general WD method to measure the exact nSTL of an acoustical element without any assumptions on its symmetry and on the termination. Similar to general TM methods for non-symmetrical elements, four microphones and two loads will be required. As a first validation of the method, symmetrical and non-symmetrical porous materials are investigated. Results are discussed and compared with some existing methods and with the classical two-load method. A perfect agreement is found with the classical two-load method.
 [DOI: 10.1121/1.3081504]

PACS number(s): 43.55.Rg, 43.58.Bh, 43.55.Ev, 43.20.Mv [KVH] Pages: 2083–2090

I. INTRODUCTION

In the past, numbers of methods have been published to measure the normal incidence plane wave sound transmission loss (nSTL) of acoustical elements (i.e. materials, mufflers, expansion chambers, etc.). Some are limited to symmetrical elements, and some are not. Since reciprocity of sound transmission has been proven[1] (at least for linear elements separating identical fluid media), the only assumption that remains is on the symmetry of the element. Since symmetrical elements rarely show a perfect symmetry in terms of its acoustical surface properties[2] (i.e., absorption, reflection or impedance), one would prefer to use a general method which does not assume symmetry.

Two types of general methods can be found in the literature. The first one is based on the transfer matrix (TM) approach, and the second one is based on the wavefield decomposition (WD) theory. A review of some of these methods is given by Tao and Seybert.[3] Following the TM approach, the acoustical element is modeled as a TM. Then, the objective of the approach is to propose an experimental procedure to determine the coefficients of the matrix from which the nSTL will be deduced. On the other hand, the WD theory deals with the complex wavefields upstream and downstream the element. The objective is to conveniently decompose these wavefields in terms of forward and backward propagating waves. From this decomposition, the nSTL can be directly deduced.

The most referred general TM methods are the two-load method proposed by Lung and Doige[4] in 1983, and the two-source location method proposed by Munjal and Doige[5] in 1990. These methods are also referred as four-microphone methods since two microphones are flush mounted on the upstream tube, and two microphones are flush mounted on the downstream tube—the acoustical element being placed between the two tubes. With the two-load method, a loudspeaker is placed at one end (upstream) to generate noise, while sound pressures are measured at the four microphones for two different downstream tube terminations (for example, anechoic, rigid, or open end). With the two-source location method, a unique arbitrary termination is used; however, this time the sound pressure measurements are performed with the loudspeaker mounted at one end, and repeated when the loudspeaker is mounted at the other end. Both methods theoretically give identical results and yield the exact nSTL of the acoustical element. An application of the TM method for the characterization of porous materials using a modified impedance tube was presented by Song and Bolton[6] for symmetrical materials in 2000, and was extended to the evaluation of the nSTL of non-symmetrical materials by Olivieri *et al.* in 2006.[7] However, it is worth mentioning that the latter method is not different to the original two-load method. In fact, Eq. (7) in the paper by Olivieri *et al.*[7] is identical (however, written differently) to Eq. (2) in the original paper by Lung and Doige.[4]

For the WD theory, the most known method was used by Seybert and Ross.[8] This method requires two flush mounted microphones on the upstream tube, and one on the downstream tube. A loudspeaker is placed at the upstream end, and an anechoic termination is assumed at the downstream end. With the two upstream microphones, one can decompose the wavefield upstream the element in terms of an incident wave (A) and a reflected wave (B). With the anechoic termination assumption, the downstream microphone di-

[a]Electronic mail: yacoubou.salissou@usherbrooke.ca

rectly measures the sound pressure of the transmitted wave (C). From the ratio (C/A) between the transmitted and incident pressures, the nSTL can be deduced. The main limitation of the method is due to the anechoic termination. In fact, constructing a fully anechoic termination is difficult. Consequently, a part of the transmitted wave is reflected by the termination, and multiple forward and backward reflections occur in the downstream tube. To attempt eliminating this problem, Chung and Blaser[9,10] also used a decomposition in the downstream tube using two flush mounted microphones. The decomposition yields one wave (C) propagating toward the termination and one wave (D) propagating back to the element to test. Again, from ratio C/A, they deduced the nSTL. This approach is not exact since wave C is not equal to the first transmitted wave. In fact, C is equal to the sum of the first transmitted wave (the only one existing in the case of an anechoic termination) and all the multiple back reflected waves propagating in the direction of the termination. In this case, the only situation where ratio C/A yields the exact nSTL is again when the termination is fully anechoic. If the termination is not fully anechoic and the back surface of the element to test is relatively absorbing (i.e., the dominating waves in the downstream tube are only the first transmitted wave and its first reflection on the termination), the Chung and Blaser four-microphone method gives a better estimation than the Ross and Seybert three-microphone method which assumes a fully anechoic termination. A downstream description similar to the one used by Chung and Blaser was also applied to the TM approach by Bonfiglio and Pompoli.[11] In their hybrid WD-TM method, only a single four-microphone measurement is required (i.e., the method only requires one load). Even if the symmetry assumption is used, their method compares well with the original two-load method for non-symmetrical elements. As previously mentioned, the method should work until the back surface of the element to test is relatively absorbing (this point will be shown in this paper). Finally, it is worth mentioning that some previous studies in the literature also showed that only the two-load methods can provide reliable value of nSTL.[12]

From this brief literature review, if the termination is not fully anechoic and the acoustical element is not symmetrical, or not fully symmetrical, the general four-microphone TM methods (two-load or two-source location) are the only methods yielding the exact nSTL. For this general situation, there exists no equivalent WD method yielding the exact nSTL. It is worth mentioning that recent attempts to generalize the WD method were published;[13,14] however, at the theoretical level, some discussions remain (they will be underlined later on in this paper).

The objective of this paper is to develop a general WD method to measure the exact nSTL of an acoustical element without any assumptions on its symmetry and on the termination. Similar to the general TM methods for non-symmetrical elements, four microphones and two loads will be required. The challenge is to use a proper theoretical description to finally retrieve the exact nSTL.

This paper is structured as follows. Section II presents the theoretical developments behind the proposed wave decomposition method. Section III tests the method using a transmission tube on symmetrical and non-symmetrical materials, and studies the effects of the terminations. Also, it compares the proposed method to other existing methods. Finally, Sec. IV draws conclusions on this work.

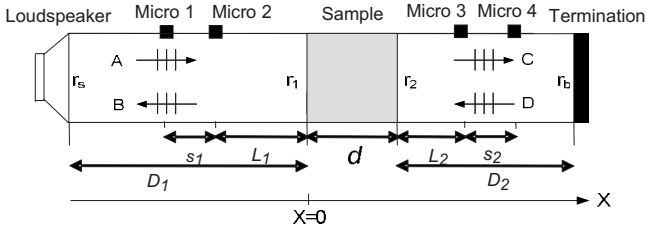


FIG. 1. Schematic view of the experimental setup.

II. THEORY

A schematic view of the impedance tube used to measure the nSTL is shown in Fig. 1. This apparatus consists of a finite-length rigid walled impedance tube with circularly shaped and uniform inner cross-section. At one end, the tube features a loudspeaker, and at the other end, the tube can be fitted with an arbitrary termination condition. The loudspeaker is used to generate a plane wave field in the impedance tube. There are two microphones flush mounted upstream the sample and two microphones flush mounted downstream the sample.

For this setup, Ho *et al.*[13] proposed an improved WD method which takes explicitly into account the multiple wave reflections in the upstream and downstream tubes. The method was limited to symmetrical acoustic systems. Using the symmetry property, Ho *et al.* assumed that the reflection coefficient (r_1) of the front surface of the sample is equal to the one of its back surface (r_2). This assumption is too restrictive since the reflection coefficient is a global surface indicator which also depends on the condition behind the sample. For instance, on the upstream side there is an air cavity with a loudspeaker, and on the downstream side there is an air cavity with an arbitrary termination. Therefore, even though the material is symmetrical, the reflection coefficient of an incident plane wave impinging its front face is different from the one of an incident plane wave impinging its back face (i.e., $r_1 \neq r_2$). A similar restrictive assumption seems to be used also by Dong-Li *et al.*[14] [for example, see Eqs. (5) and (10) in Ref. 14].

Based on the work by Ho *et al.*,[13] under steady state harmonic condition, and eventually making no assumption on the reflection coefficients, the total wave propagating toward the sample in the upstream tube can be explicitly written as

$$p_u^a(x) = ae^{-jk_0x}(1 + r_1r_se^{2jk_0D_1} + (r_1r_s)^2e^{4jk_0D_1} + \cdots)$$
$$= \frac{ae^{-jk_0x}}{1 - r_1r_se^{2jk_0D_1}}. \quad (1)$$

The first term in Eq. (1) is the wave just emitted by the loudspeaker, the second term is the wave reflected by the sample and the loudspeaker while traveling the entire length of the front tube twice, the third term is the second wave

reflected by the sample and the loudspeaker while traveling the entire length of the front tube four times, and so on. The final result is the sum of an infinite number of these multiple reflections

Similarly, the total wave traveling away from the sample in the upstream tube can be explicitly written as

$$p_u^b(x) = ar_1 e^{jk_0 x}(1 + r_1 r_s e^{2jk_0 D_1} + (r_1 r_s)^2 e^{4jk_0 D_1} + \cdots) = \frac{ar_1 e^{jk_0 x}}{1 - r_1 r_s e^{2jk_0 D_1}}. \quad (2)$$

Here, it is worth commenting on the nature of the steady state harmonic condition. Under this steady state condition, if one tries interpreting the first term in Eq. (2), it has to be seen as the first incident wave reflected by the sample which is backed by the downstream tube and its termination. With two microphones flush mounted on the upstream tube, this reflection coefficient r_1 is readily obtained using the transfer function method [ASTM E1050-98 (Ref. 15) or ISO 10534-2:1998 (Ref. 16)]. Consequently, this first reflected term implicitly takes into account all the multiple reflections of the first incident wave taking place in the sample and in the downstream tube. This would not be true, if the observations were done in the time domain. The aforementioned comment is also valid for all the terms in Eqs. (1) and (2), and for all the equations below.

Adding up Eqs. (1) and (2), the total sound pressure field in the upstream tube is given by

$$p_u(x) = \frac{a}{1 - r_1 r_s e^{2jk_0 D_1}}(e^{-jk_0 x} + r_1 e^{jk_0 x}). \quad (3)$$

With the origin fixed at the front surface of the sample, one easily finds that the total incident sound pressure at $x=0$ is $a/(1-r_1 r_s \exp(2jk_0 D_1))$. This incident wave is attenuated as it goes through the sample by the complex sound transmission coefficient t. On the rear face, the incident pressure is now $[a/(1-r_s r_1 \exp(2jk_0 D_1))]t$. Treating this term as the sound source in the downstream tube, and following the same approach as in the upstream tube, the total pressure field in the downstream tube is

$$p_d(x) = \frac{a}{(1 - r_1 r_s e^{2jk_0 D_1})(1 - r_2 r_b e^{2jk_0 D_2})}(e^{-jk_0 x} + r_b e^{jk_0 x})t. \quad (4)$$

Contrary to Ho *et al.*,[13] the reflection coefficient r_2 in Eq. (4) is not assumed equal to r_1. In fact, r_2 is the reflection coefficient seen by x-negative propagating waves in the downstream tube, and r_1 is the reflection coefficient seen by x-positive propagating waves in the upstream tube. Since D_1 is not necessarily equal to D_2, and the reflection coefficient of the loudspeaker (r_s) is not equal to the one of the termination (r_b), it is clear that r_1 and r_2 will never be equal, even though the sample is symmetrical.

Equations (3) and (4) are now used to evaluate the sound pressures at microphones 1–4. They are, respectively, given by

$$p_1 = \frac{a}{1 - r_1 r_s e^{2jk_0 D_1}}(e^{jk_0(L_1+s_1)} + r_1 e^{-jk_0(L_1+s_1)}), \quad (5)$$

$$p_2 = \frac{a}{1 - r_1 r_s e^{2jk_0 D_1}}(e^{jk_0 L_1} + r_1 e^{-jk_0 L_1}), \quad (6)$$

$$p_3 = \frac{a}{(1 - r_1 r_s e^{2jk_0 D_1})(1 - r_2 r_b e^{2jk_0 D_2})}(e^{-jk_0 L_2} + r_b e^{jk_0 L_2})t, \quad (7)$$

$$p_4 = \frac{a}{(1 - r_1 r_s e^{2jk_0 D_1})(1 - r_2 r_b e^{2jk_0 D_2})} \times (e^{-jk_0(L_2+s_2)} + r_b e^{jk_0(L_2+s_2)})t. \quad (8)$$

Defining the transfer function between microphones 1 and 2 by $H_{12}=p_1/p_2$, and substituting p_1 and p_2 with Eqs. (5) and (6), respectively, the reflection coefficient of the sample on the loudspeaker side can be expressed in terms of the transfer function as

$$r_1 = \frac{H_{12} - e^{jk_0 s_1}}{e^{-jk_0 s_1} - H_{12}} e^{2jk_0 L_1}. \quad (9)$$

Similarly, defining the transfer function $H_{43}=p_4/p_3$ and using Eqs. (7) and (8), the reflection coefficient of the termination can be expressed as

$$r_b = \frac{H_{43} - e^{-jk_0 s_2}}{e^{jk_0 s_2} - H_{43}} e^{-2jk_0 L_1}. \quad (10)$$

Finally, from Eqs. (6) and (7), the transfer function between microphones 3 and 2 is given by

$$H_{32} = \frac{p_3}{p_2} = \frac{1}{1 - r_2 r_b e^{2jk_0 D_2}} \frac{e^{-jk_0 L_2} + r_b e^{jk_0 L_2}}{e^{jk_0 L_1} + r_1 e^{-jk_0 L_1}} t. \quad (11)$$

In this equation, one can note that the transmission coefficient t and reflection coefficient r_2 are the only remaining two unknowns. If one needs to determine the transmission coefficient, the reflection coefficient needs first to be found. The reflection coefficient r_2 can be obtained by changing the termination of the tube (second load). Indeed, using a termination with a different reflection coefficient r_b' will change the reflection coefficient of the sample on the loudspeaker side to r_1'. However, since the upstream tube remains unaffected by this operation, the reflection coefficient r_2 remains unchanged. For the second load, the modified reflection coefficients are

$$r_1' = \frac{H_{12}' - e^{jk_0 s_1}}{e^{-jk_0 s_1} - H_{12}'} e^{2jk_0 L_1}, \quad (12)$$

and

$$r_b' = \frac{H_{43}' - e^{-jk_0 s_2}}{e^{jk_0 s_2} - H_{43}'} e^{-2jk_0 L_2}, \quad (13)$$

where H_{12}' and H_{43}' are, respectively, the new transfer functions between microphones 1 and 2 ($H_{12}'=p_1'/p_2'$), and microphones 4 and 3 ($H_{43}'=p_4'/p_3'$) under the second load condition. For this second load, the new transfer function between microphones 3 and 2 is given by

$$H'_{32} = \frac{p'_3}{p'_2} = \frac{1}{1 - r_2 r'_b e^{2jk_0 D_2}} \frac{e^{-jk_0 L_2} + r'_b e^{jk_0 L_2}}{e^{jk_0 L_1} + r'_1 e^{-jk_0 L_1}} t. \quad (14)$$

Introducing the parameters

$$h = H_{32}/H'_{32}, \quad (15)$$

and

$$r = \frac{e^{-jk_0 L_2} + r_b e^{jk_0 L_2}}{e^{-jk_0 L_2} + r'_b e^{jk_0 L_2}} \frac{e^{jk_0 L_1} + r'_1 e^{-jk_0 L_1}}{e^{jk_0 L_1} + r_1 e^{-jk_0 L_1}}, \quad (16)$$

one can deduce from Eqs. (11) and (14) the reflection coefficient of the sample on the downstream side

$$r_2 = \frac{r - h}{r r'_b - h r_b} e^{-2jk_0 D_2}. \quad (17)$$

Finally, using Eqs. (17) and (11), the transmission coefficient and the nSTL of the sample can be expressed, respectively, as

$$t = H_{32}(1 - r_2 r_b e^{2jk_0 D_2}) \frac{e^{jk_0 L_1} + r_1 e^{-jk_0 L_1}}{e^{-jk_0 L_2} + r_b e^{jk_0 L_2}}. \quad (18)$$

and

$$\text{nSTL} = -20 \log(t). \quad (19)$$

III. EXPERIMENTAL TESTS

The six transfer functions (H_{12}, H_{43}, H_{32}, H'_{12}, H'_{43}, and H'_{32}) are measured similarly to the standard transfer function method[15] using a homemade impedance tube. The diameter of the tube is 44.5 mm and its validity range is from 300 to 4200 Hz. The other parameters of the setup shown in Fig. 1 are $D_1 = 190$ mm, $D_2 + d = 362$ mm, $s_1 = 25$ mm, $s_2 = 25$ mm, $L_1 = 45$ mm, and $L_2 + d = 271$ mm. A dual-channel fast Fourier transform analyzer and two 1/4 in. microphones are used to conduct measurements successively in pairs (1-2, 3-2, 4-3) at the four locations. Implementation of a four-microphone measurement is also possible to reduce measurement time. Random noise using a Hanning window and an averaging over 100 measurements are used. The standard sensor switching technique was used to compensate the microphone phase and amplitude mismatch.

Two symmetrical and two non-symmetrical materials are used for these experimental tests. For each sample, four different termination combinations are tested to determine the nSTL: rigid-anechoic, rigid-absorbent, absorbent-absorbent, and anechoic-absorbent. Three different absorbents are used. Absorbent 1 is a 22-mm-thick floor material made of porous recycled foams backed by a rigid end cap. Absorbent 2 is a 15-mm-thick car headliner foam backed by a rigid end cap. Absorbent 3 is a 25-mm-thick melamine foam backed by a rigid end cap. The rigid termination is the rigid end cap alone. The anechoic termination is a 200-mm conical block of melamine foam backed by the rigid end cap. The measured reflection coefficient amplitudes of the five possible terminations are shown in Fig. 2. The reflection coefficient of the rigid termination is 97%–99% instead of its ideal value of 100%. This deviation may be attributed to some energy dissipation in the viscous and thermal skin layers of the tube, and to uncertainties on the exact location of the acoustic centers of the microphones. The reflection of the absorbing ("anechoic") termination ranges from 70% at low frequencies to 5% at high frequencies. This represents a significant deviation from its ideal value of 0%. Finally, one can note that the reflection coefficients of the two Absorbents 1 and 2 are close for frequencies lower than 2000 Hz.

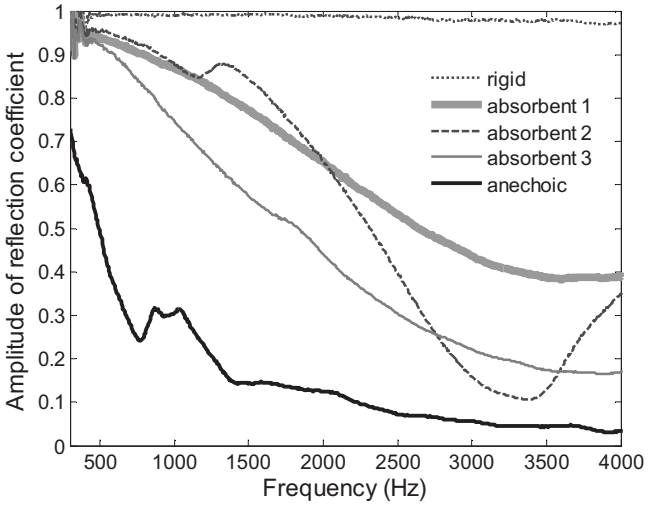


FIG. 2. Amplitudes of the reflection coefficient of the five termination loads which can be used in the measurements of the normal incidence sound transmission.

A. Symmetrical materials

The first tested symmetrical material is a 25-mm-thick melamine foam. The second symmetrical material is a 15-mm-thick car headliner. The relative difference curves on the amplitude of the surface impedance of each of those materials are plotted in Fig. 3. From these curves and according to the symmetry criterion presented elsewhere by the authors,[2] both materials can be considered symmetrical in terms of their acoustical properties.

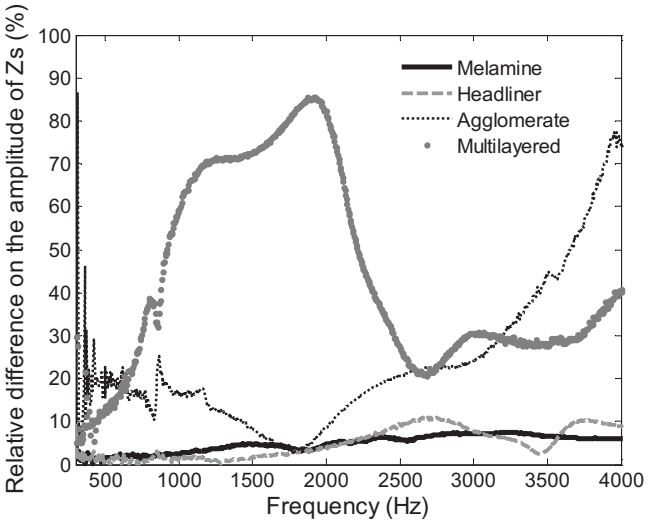


FIG. 3. Relative difference on the amplitude of the surface impedance (Z_s) of the tested samples.

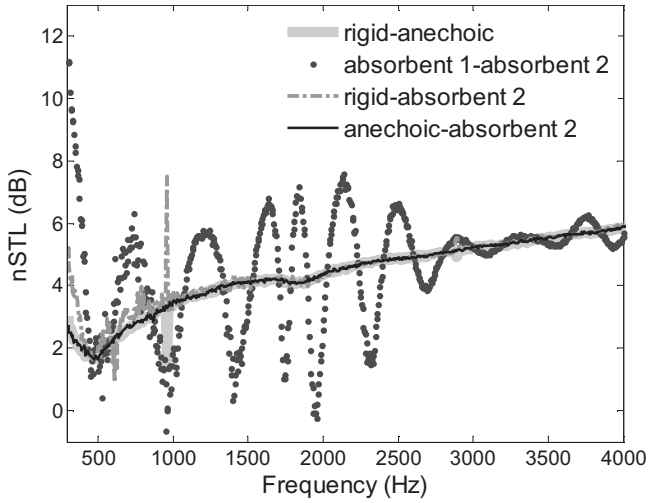


FIG. 4. nSTL of a 25-mm-thick melamine foam measured using the proposed WD method with four different load combinations.

1. Melamine foam

Using the developed method, the measured nSTLs of the melamine foam for four combinations of two loads are shown in Fig. 4. Note that the absorbing loads used in this characterization are absorbents 1 and 2. As one can note, all combinations of two loads provide similar results, except for the combination of the two absorbents. Indeed, in order to have a better result, one needs to choose a combination of two loads with very different reflection properties. This condition is only achieved with the three other load combinations. Also, for the combinations using the rigid end, one can note a singularity around 1000 Hz corresponding to the resonance frequency. To avoid this resonance, a termination combination made of anechoic-absorbent or two absorbents with different reflection coefficients should be chosen.

Next, the results obtained from the proposed method and those obtained from several existing four-microphone methods (see Ref. 17) are compared in Fig. 5. For the methods using two loads, the load combination absorbent 1-anechoic was used. For the methods using one load, the anechoic load was used. Since all the tested methods use the same four microphones, the same set of pressure measurements were used to feed these methods. Consequently, equivalent methods should yield identical results.

In Fig. 5(a), one can note that the proposed two-load WD method and the classical two-load TM method by Lung and Doige[4] yield the same (identical) results. This proves that the aforementioned theoretical development seems general as expected. In contrast, the one-load WD method by Ho *et al.*[13] appears to be inaccurate [see Fig. 5(b)]. Indeed, the result from this method exhibits a periodic fluctuation. The fluctuation is due to the fact that Ho *et al.* assumed that $r_1 = r_2$ for a symmetrical material. As pointed out previously, even though the material is acoustically symmetrical, r_1 is not necessarily equal to r_2 since the conditions upstream and downstream the sample are different. Note that this error will increase as the amount of incident power backward the sample increases. This justifies the need of an anechoic end in order to apply this approximated method. At high frequencies (>3000 Hz), the partially anechoic end used in this work almost behaves as a fully anechoic termination, with a reflection coefficient amplitude less than 0.1 (see Fig. 2). As a result, the fluctuation almost disappears in this frequency range and the approximated nSTL overlaps with the one obtained from the proposed general WD method.

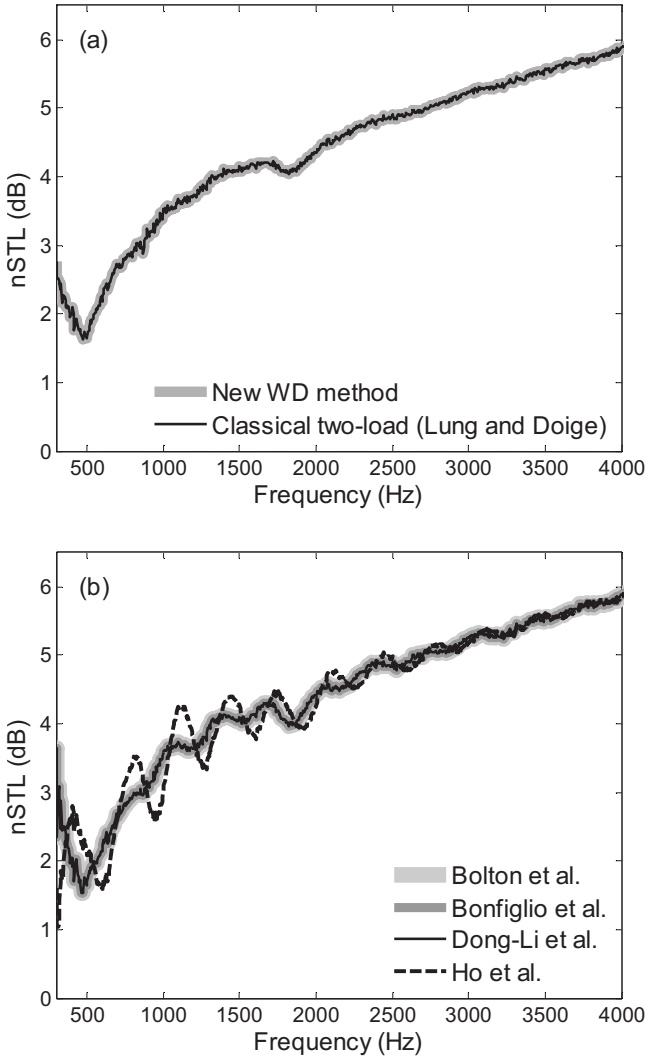


FIG. 5. Measured nSTL of a 25-mm-thick melamine foam. Comparison between the different four-microphone methods. (a) Two-load methods. (b) One-load methods. The two-load methods use the absorbent 1-anechoic load combination. The one-load methods use the relatively anechoic load.

The TM method by Song and Bolton,[6] the hybrid WD-TM method by Bonfiglio and Pompoli,[11] and the WD method by Dong-Li *et al.*[14] are compared in Fig. 5(b). All three methods are one-load methods and yield similar (identical) results. This is logical since the material is absorbing and symmetrical, and all three methods take into account (but differently) the transmitted and transmitted-reflected waves in the downstream segment. However, compared to the results obtained with the proposed method and the classical two-load method in Fig. 5(a), the one-load methods show slight fluctuations. These fluctuations are mainly due to the fact that higher order reflected waves in the downstream tube are not considered by these one-load methods, and/or

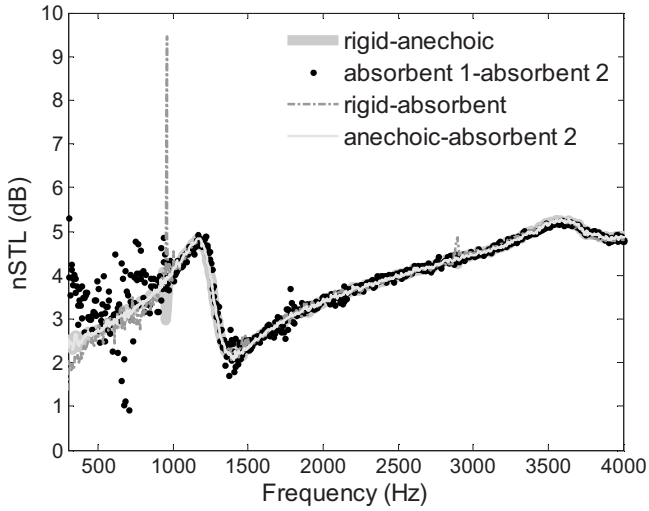


FIG. 6. nSTL of a 15-mm-thick car headliner foam measured using the proposed WD method with four different load combinations.

the material is not fully symmetrical (see Fig. 3). However, those fluctuations are minor compared to those obtained with the WD method by Ho *et al.*[13]

2. Headliner foam

For this sample, the absorbing loads used in the nSTL characterizations are absorbents 1 and 3. The measured nSTLs for the headliner are shown in Fig. 6. The same comments on the effects of the termination combination can be done. Again, the results are in good agreement. For the load combinations using the rigid end, the singularity around 1000 Hz is still observed. Because of the more significant difference between the reflection coefficients of the two absorbents used in combination, the corresponding result is more accurate. However, one can note a relative dispersion at low frequencies. This corresponds to the range in which the reflection coefficients of the two absorbents are close. In Fig. 7, the result from the proposed method is compared to those obtain from the other two-load and one-load methods. Analyzing those results yield exactly the same conclusions previously obtained.

B. Nonsymmetrical materials

The first tested non-symmetrical sample, referenced as multilayer 1, is a 20-mm-thick multilayer made of a thick layer of rigid closed-cell polyurethane foam in sandwich between two thin impervious paper films. The second non-symmetrical sample, referenced as multilayer 2, is a 20-mm-thick multilayer made of a thick layer of rigid closed-cell polyurethane foam in sandwich between one thin paper film and one thin impervious aluminum film. The relative difference curves on the amplitude of the surface impedance of each of these multilayered panels are also plotted in Fig. 3. As expected, according to the symmetry criterion presented elsewhere by the authors,[2] these multilayered samples are clearly non-symmetrical in terms of their acoustical properties, mostly the second multilayer. Their nSTLs are measured using the termination combination made of absorbent 1 and the partially anechoic end. The results are plotted in Fig.

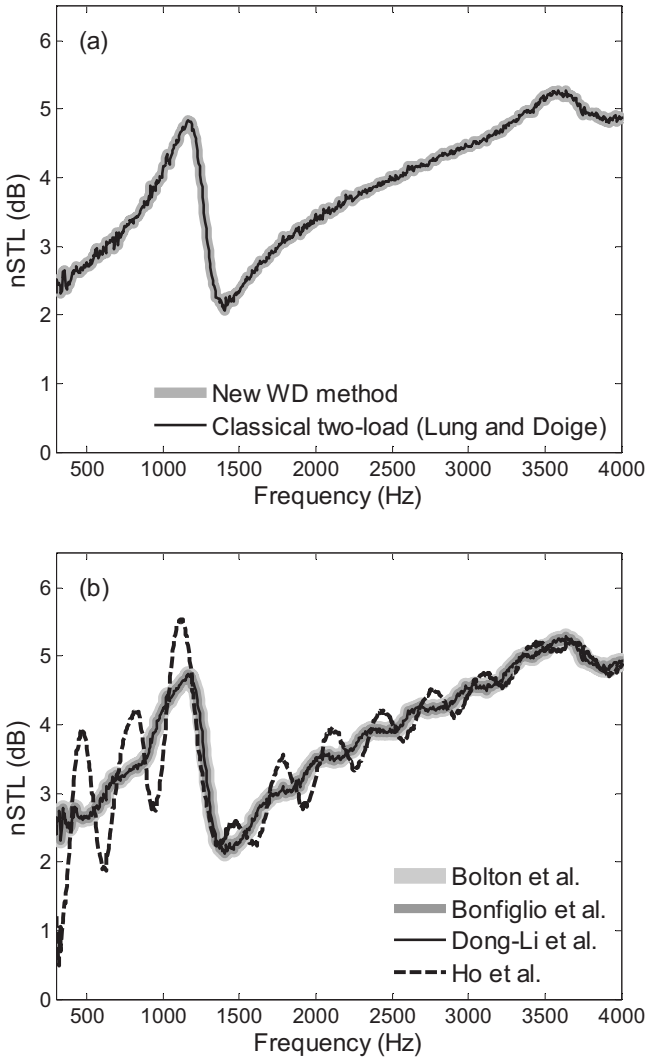


FIG. 7. Measured nSTL of a 15-mm-thick car headliner foam. Comparison between the different four-microphone methods. (a) Two-load methods. (b) One-load methods. The two-load methods use the absorbent 1-anechoic load combination. The one-load methods use the relatively anechoic load.

8. Again, both the proposed method and the classical two-load method[4] yield similar results for both asymmetrical multilayers [see Figs. 8(a) and 8(c)]. However, for all the one-load methods, the fluctuations are more important than those observed in Fig. 5 for the symmetrical materials. The fluctuations are of similar magnitudes for both multilayers, even if multilayer 2 is much more asymmetrical than multilayer 1. This proves the fluctuations are mostly attributed to the higher order reflections in the downstream tube which are not correctly modeled with the one-load methods. These higher order reflections are more important since the back surfaces of the multilayers are reflective and not absorbing as for the melamine and headliner foams.

IV. CONCLUSION

In this paper, based on the work by Ho *et al.*,[13] a general two-load WD method to measure the nSTL of an acoustical element without any assumptions on its symmetry and on the termination has been presented and discussed. Comparisons with the classical two-load TM method by Lung and Doige[4]

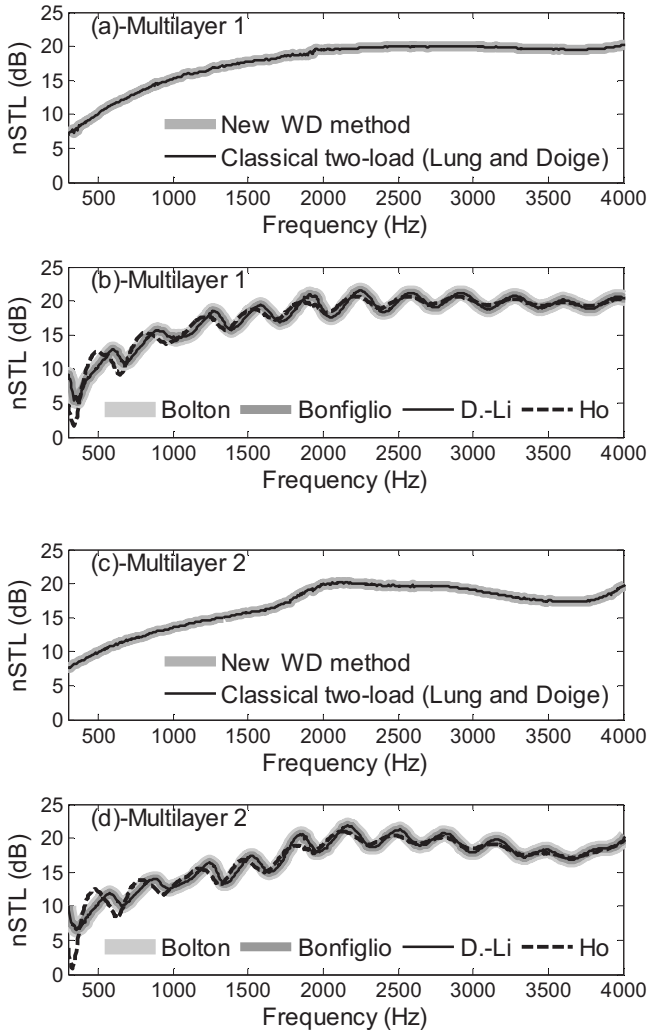


FIG. 8. Measured nSTL of the two asymmetrical multilayers. Comparison between the developed new WD method and the classical two-load method for (a) multilayer 1 and (c) multilayer 2. Comparison between the one-load methods for (b) multilayer 1 and (d) multilayer 2.

showed that the proposed method yields identical results. With a view to obtain accurate results, it was shown that one needs to use two loads of relatively different reflection coefficients. The ideal load combination can be made from a partially anechoic termination and an absorbing material, or from two absorbing materials with different reflective properties.

Also, it was shown that the recent one-load methods yield similar results, and may yield to fluctuating transmission loss curves when applied to asymmetrical acoustical elements having a reflective back surface. The fluctuation is less important if the back surface of the elements is absorbing in view to minimize the effects of the higher order reflections in the downstream tube.

Finally, it is worth emphasizing that the determination of the nSTL of an arbitrary acoustical element requires necessarily (at least at this time and until evidence of the contrary) four microphones and two loads (or two-source locations). The one-load methods remain approximated methods which can work well in many cases; however, since all cases are not tested yet, these methods must be used with cautions. Their main advantage is that they only use one set of measurements, making their implementation twice less time consuming than the general four-microphone methods.

ACKNOWLEDGMENTS

This work was supported in part by Grants-in-Aid from Alcan and NSERC A part of the research presented in this paper was also supported by the Fond québécois de la recherche sur la nature et les technologies (F.Q.R.N.T.) by the intermediary of the Aluminium Research Centre-REGAL.

NOMENCLATURE

s_1 = distance between microphones 1 and 2
s_2 = distance between microphones 3 and 4
L_1 = distance between microphone 2 and the sample on the source side
L_2 = distance between microphone 3 and the sample on the termination side
D_1 = distance between the loudspeaker and the sample
D_2 = distance between the termination and the sample
d = thickness of the sample
r_s = complex reflection coefficient at the surface of the loudspeaker
r_b = complex reflection coefficient at the surface of the termination
r_1 = complex reflection coefficient at the surface of the sample on the source side
r_2 = complex reflection coefficient at the surface of the sample on the termination side
c = speed of sound in air
f = frequency
$k_0 = 2\pi f/c$ = wave number in air
t = complex transmission coefficient of the sample
x = coordinate along the tube length
a = amplitude of the wave emitted by the loudspeaker
$p_u(x)$ = sound pressure at position x upstream the sample
$p_d(x)$ = sound pressure at position x downstream the sample
p_i = sound pressure at microphone i
p_i' = sound pressure at microphone i under a different tube end condition

[1]J.-F. Allard, B. Brouard, D. Lafarge, and W. Lauriks, "Reciprocity and antireciprocity in sound transmission through layered materials including elastic and porous media," Wave Motion **17**, 329–335 (1993).
[2]Y. Salissou and R. Panneton, "Quantifying the through-thickness asymmetry of sound absorbing porous materials," J. Acoust. Soc. Am. **124**, EL28–EL33 (2008).
[3]Z. Tao and A. F. Seybert, "A review of current techniques for measuring muffler transmission loss," SAE Technical Paper No. 2003-01-1653 (2003).
[4]T. Y. Lung and A. G. Doige, "A time-averaging transient testing method for acoustic properties of piping systems and mufflers," J. Acoust. Soc. Am. **73**, 867–876 (1983).
[5]M. L. Munjal and A. G. Doige, "Theory of a two source-location method

for direct experimental evaluation of the four-pole parameters of an aeroacoustic element," J. Sound Vib. **141**(2), 323–333 (1990).
[6]B. H. Song and J. S. Bolton, "A transfer matrix approach for estimating the characteristic impedance and wave numbers of limp and rigid porous materials," J. Acoust. Soc. Am. **107**, 1131–1152 (2000).
[7]O. Olivieri, J. S. Bolton, and T. Yoo, "Measurement of transmission loss of materials using a standing wave tube," Proceedings of Inter-Noise 2006, Honolulu, HI, (2006).
[8]A. F. Seybert and D. F. Ross, "Experimental determination of acoustic properties using a two-microphone random-excitation technique," J. Acoust. Soc. Am. **61**, 1362–1370 (1977).
[9]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. I. Theory," J. Acoust. Soc. Am. **68**, 907–913 (1980).
[10]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. II. Experiment," J. Acoust. Soc. Am. **68**, 914–921 (1980).
[11]P. Bonfiglio and F. Pompoli, "A single measurement approach for determination of the normal incidence sound transmission loss," J. Acoust. Soc. Am. **124**, 1577–1583 (2008).
[12]G. Pispola, K. V. Horoshenkov, and F. Asdrubali, "Transmission loss measurement of consolidated granular media (L)," J. Acoust. Soc. Am. **117**, 2716–2719 (2005).
[13]K. M. Ho, Z. Yang, X. X. Zhang, and P. Sheng, "Measurement of sound transmission through panels of locally resonant materials between impedance tubes," Appl. Acoust. **66**, 751–765 (2005).
[14]P. Dong-Li, H. Peng, and Z. Bei-Li, "The modified method of measuring the complex transmission coefficient of acoustical panel in impedance tube," Appl. Acoust. **69**, 1240–1248 (2008).
[15]"Standard test method for impedance and absorption of acoustical materials using a tube, two microphones, and a digital frequency analysis system," ASTM Standard No. E1050-98 (ASTM International, West Conshohocken, PA, 2006).
[16]"Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes—Part 2: Transfer-function method," ISO Standard 10534-2:1998 (International Organization for Standardization, Geneva, 1998).
[17]Note that the method of Ho *et al.* (Ref. 13) uses only three microphones since the reflection coefficient of the termination r_b is assumed to be measured beforehand. In our application of the method of Ho *et al.*, microphones 3 and 4 were used to evaluate r_b.

A realization of sound focused personal audio system using acoustic contrast control

Ji-Ho Chang,[a] Chan-Hui Lee,[b] Jin-Young Park,[c] and Yang-Hann Kim[d]
Department of Mechanical Engineering, Center for Noise and Vibration Control, Korea Advanced Institute of Science and Technology, Science Town, Daejeon 305-701, Korea

(Received 6 May 2008; revised 21 January 2009; accepted 24 January 2009)

A personal audio system that does not use earphone or any wire would have great interest and potential impact on the audio industries. In this study, a line array speaker system is used to localize sound in the listening zone. The contrast control [Choi, J.-W. and Kim, Y.-H. (2002). J. Acoust. Soc. Am. **111**, 1695–1700] is applied, which is a method to make acoustically bright zone around the user and acoustically dark zone in other regions by maximizing the ratio of acoustic potential energy density between the bright and the dark zone. This ratio is regarded as acoustic contrast, analogous with what is used for optical devices. For the evaluation of the performance of acoustic contrast control, experiments are performed and the results are compared with those of uncontrolled case and time reversal array. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3082114]

PACS number(s): 43.60.Fg, 43.38.Hz [AJZ] Pages: 2091–2097

I. INTRODUCTION

Recently personal devices such as desktops, laptop computers, portable multimedia player, and cellular phones are widely used. The users listen to sound from the devices by earphones or loudspeakers, but it induces some troubles; earphones are not only burdensome to wear but also uncomfortable to use for a long time, and loudspeakers make sound that others do not want to hear.

A personal audio system that uses loudspeakers but focuses sound only around the user can solve these problems of earphones and loudspeakers. In this study, we propose a system to localize sound only around the user by using a line array of control sources. That is, the objective of this study is to increase acoustic potential energy only in a zone including the user but to reduce in the other zone by controlling the input signals, which are fed into control sources. In order to realize it, we apply the previous research on acoustic contrast control.[1]

There are other sound focusing methods such as parametric array, time reversal array, beamforming, and inverse filtering (for example, see Refs. 2–5). They can be good candidates to tackle this problem, but they have limitations; for example, parametric array consumes too much power to be applied to portable devices and occupies too large area.

In Sec. II, the acoustic contrast control is briefly explained. In Secs. III and IV, the process that the acoustic contrast control is applied to the system is described regardless of the size of the array. Then in Sec. V, a specific example of the system that can be attached on 17 in. monitor display is represented. In order to verify the system, we have experiments and compare the results with those of uncontrolled case and time reversal array to evaluate the performance of acoustic contrast control.

II. THEORETICAL BACKGROUND: ACOUSTIC CONTRAST CONTROL

"Acoustic contrast control" has been proposed by Choi and Kim.[1] "Acoustic contrast" means the ratio of acoustic potential energy density between acoustically bright zone and dark zone that we selected (Fig. 1). Acoustic contrast control provides us the optimal solution set that maximizes acoustic contrast. By applying the method, we can increase or reduce acoustic potential energy density as we want.

Let us consider L control sources located at the positions $\boldsymbol{r}_s=[\boldsymbol{r}_s^{(1)}\boldsymbol{r}_s^{(2)}\cdots\boldsymbol{r}_s^{(L)}]$ and denote source inputs as $\mathbf{q}(\omega)=[q^{(1)}(\omega)q^{(2)}(\omega)\cdots q^{(L)}(\omega)]^T$ where ω indicates radial frequency. The transfer functions between source inputs and pressure at a position $\boldsymbol{r}$ can be expressed by $\mathbf{h}(\boldsymbol{r}|\boldsymbol{r}_s,\omega)=[h(\boldsymbol{r}|\boldsymbol{r}_s^{(1)},\omega)h(\boldsymbol{r}|\boldsymbol{r}_s^{(2)},\omega)\cdots h(\boldsymbol{r}|\boldsymbol{r}_s^{(L)},\omega)]$. A sound field generated by the control sources can be expressed as

$$p(\boldsymbol{r},\omega)=\mathbf{h}(\boldsymbol{r}|\boldsymbol{r}_s,\omega)\mathbf{q}(\omega). \tag{1}$$

Because the acoustic potential energy density $e(\omega)$ in zone V is proportional to square of pressure, we can write it as

$$\begin{aligned} e(\omega)&=\frac{1}{V}\int_V p(\boldsymbol{r},\omega)^H p(\boldsymbol{r},\omega)dV \\ &=\mathbf{q}(\omega)^H\left(\frac{1}{V}\int_V \mathbf{h}(\boldsymbol{r}|\mathbf{r}_s,\omega)^H\mathbf{h}(\boldsymbol{r}|\boldsymbol{r}_s,\omega)dV\right)\mathbf{q}(\omega) \\ &=\mathbf{q}(\omega)^H\mathbf{R}(\omega)\mathbf{q}(\omega), \end{aligned} \tag{2}$$

where V means the volume of control zone and $\mathbf{R}(\omega)$ is the spatial correlation matrix. If we select acoustically bright zone V_b and dark zone V_d, then the acoustic potential energy density in the bright zone $e_b(\omega)$ and that in the dark zone $e_d(\omega)$ are, respectively, expressed by

$$e_b(\omega)=\mathbf{q}(\omega)^H\mathbf{R}_b(\omega)\mathbf{q}(\omega) \tag{3}$$

and

[a]Electronic mail: chang.jiho@gmail.com
[b]Electronic mail: chance99@kaist.ac.kr
[c]Electronic mail: jypark1979@kaist.ac.kr
[d]Electronic mail: yanghannkim@kaist.ac.kr

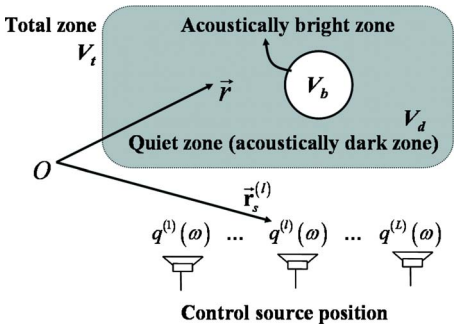


FIG. 1. (Color online) Definition of acoustic variables and control zone: V_t, total control zone ($V_t = V_b \cup V_d$); V_b, acoustically bright zone; V_d, acoustically dark zone; $\mathbf{r}_s^{(l)}$, position of the lth control source; and $q^{(l)}(\omega)$, the lth control source input.

$$e_d(\omega) = \mathbf{q}(\omega)^H \mathbf{R}_d(\omega)\mathbf{q}(\omega), \tag{4}$$

where $\mathbf{R}_b(\omega)$ and $\mathbf{R}_d(\omega)$ are the spatial correlation matrices of the bright zone and the dark zone, respectively.

Acoustic contrast $\beta(\omega)$ is defined by the ratio of the potential energy density between the bright and the dark zone as follows:

$$\beta(\omega) = \frac{e_b(\omega)}{e_d(\omega)} = \frac{\mathbf{q}(\omega)^H \mathbf{R}_b(\omega)\mathbf{q}(\omega)}{\mathbf{q}(\omega)^H \mathbf{R}_d(\omega)\mathbf{q}(\omega)}. \tag{5}$$

It is proved that the contrast control solution $\mathbf{q}_c(\omega)$ that maximizes $\beta(\omega)$ is determined to be the eigenvector which is correspondent to the maximum eigenvalue of the matrix $\mathbf{R}_d(\omega)^{-1}\mathbf{R}_b(\omega)$.[1]

III. PROBLEM DEFINITION

Before we apply the acoustic contrast control, it is noteworthy that the positions of control sources, the bright zone, and the dark zone have to be determined.

For simplicity, let us assume that sources radiate sound in free field. This is practically acceptable because the user uses the device within a short distance and direct sound is dominant compared to reverberant sound.

A. The positions of control sources

Because we have interest in a personal device that can be easily moved, the array of control sources is assumed to be attached on the device [Fig. 2 (left)]. The control sources are to be placed outside the display so that it does not obstruct the vision. The display has usually rectangular shape; therefore a line array can be regarded to be attached on the upper part or lower part of the display [Fig. 2 (left)]. Let us

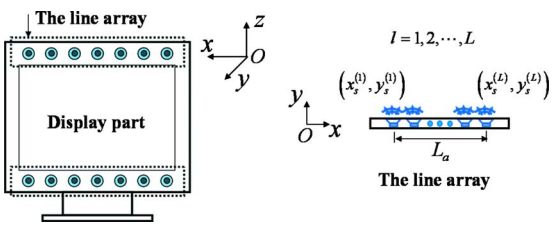


FIG. 2. (Color online) The line array of control sources that can be attached on the upper or lower side of display part (left) and the position of control sources (right).

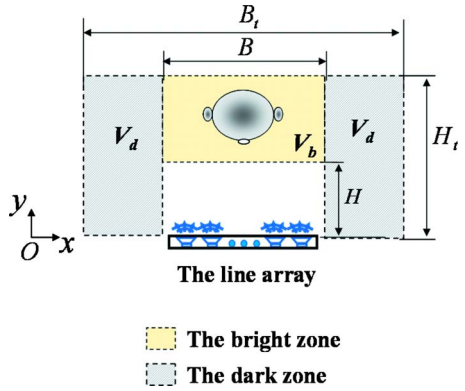


FIG. 3. (Color online) The bright zone and the dark zone: The bright zone is the region that includes the user's head in front of the array and defined as $\{(x,y)|H \le y \le H_t, -B/2 \le x \le B/2\}$, and the dark zone is the region near the bright zone defined as $\{(x,y)|-B_t/2 \le x \le -B/2$ or $B/2 \le x \le B_t/2, 0 \le y \le H_t\}$.

also assume the line array has L control sources on x axis and aperture size of the line array is L_a [Fig. 2 (right)]. Then the position of the lth control source can be expressed by

$$(x_s^{(l)}, y_s^{(l)}) = \left(-\frac{L_a}{2} + (l-1)\frac{L_a}{L-1}, 0\right), \tag{6}$$

where $l = 1,2,\ldots,L$ as illustrated in Fig. 2 (right).

B. The bright zone

The sound field generated by the line array that we assumed has dependency around the axis of the array (x axis) so that we cannot control sound field around the axis independently. Therefore we control the sound in two-dimensional plane as defined in Fig. 3. The bright zone V_b has to include the user's head. The width of the bright zone is denoted as B, and it has to be larger than the width of the user's head. This allows the head movement without inducing abrupt change in sound pressure that is designed to be generated in the bright zone.

The distance between the array and the end of the zone is denoted as H_t, as shown in Fig. 3. The distance between the array and the front of the zone is denoted as H. They determine the position and size of the user's head. It is noteworthy that H has to be big enough to avoid the effect of evanescent wave that is possible in the vicinity of the array. In other words, kH must be bigger than 1 ($kH > 1$) where k indicates the wave number.

Therefore acoustic bright zone V_b is defined as the rectangular shape region, $\{(x,y)|H \le y \le H_t, -B/2 \le x \le B/2\}$ (Fig. 3).

C. The dark zone

The other entire region except the bright zone can be regarded as dark zone V_d. However, the front region of the array, $\{(x,y)|-B/2 < x < B/2, 0 < y < H\}$, does not have to be dark zone because no one will be between the device and the user (Fig. 3). The region away from the bright zone in y direction, $H_t < y$, is not needed to be bright zone nor dark

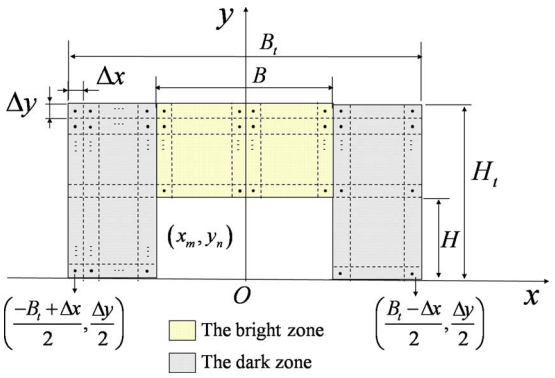


FIG. 4. (Color online) The measurement points: $(x_m, y_n)=((-B_t+\Delta x)/2+(m-1)\Delta x, \Delta y/2+(n-1)\Delta y)$, where $m=1,2,\ldots,B_t/\Delta x$ and $n=1,2,\ldots,H_t/\Delta y$.

zone because the region is out of our interest. The rear region of the array, $y<0$, is excluded because the region is out of interest as well.

The width of total zone that includes both the bright zone and the dark zone is denoted by B_t. To include the other entire zone, B_t can be regarded as infinity. However, even if B_t is a finite value, the contrast control makes the other entire region dark because pressure in the outer region of the dark zone is as low as pressure in the dark zone. Therefore, B_t can be a finite value and the dark zone can be a region that includes only those who are right beside the user.

Therefore, dark zone V_d is defined as the rectangular shaped region of $\{(x,y)|-B_t/2\leq x\leq -B/2$ or $B/2\leq x\leq B_t/2, 0\leq y\leq H_t\}$, as shown in Fig. 3.

IV. SOLUTION METHOD

In practice, the sound field of interest is discretized by measurement points, and then the acoustic contrast control is applied. The solution that maximizes the acoustic contrast is obtained at each frequency, and then the solutions at the frequencies construct a multi-channel filter to filter a monosignal.

A. Applying the acoustic contrast control at each frequency

Let us denote the measurement spacing with respect to x and y as Δx and Δy (see Fig. 4). It is noteworthy that the spacings, Δx and Δy, have to be smaller than at least half of the minimum wavelength for all frequency range to avoid spatial aliasing problem. Then the positions can be written as

$$(x_m, y_n) = \left(\frac{-B_t+\Delta x}{2} + (m-1)\Delta x, \frac{\Delta y}{2} + (n-1)\Delta y\right), \quad (7)$$

where $m=1,2,\ldots,B_t/\Delta x$ and $n=1,2,\ldots,H_t/\Delta y$. B_t and B are multiples of $2\Delta x$, and H_t and H are multiples of Δy, then m and n are integers.

The transfer functions between source inputs and pressure at the position (x_m, y_n) by the L control sources with source strength of unity can be expressed by

$$\mathbf{h}(x_m, y_n|\mathbf{x}_s, \mathbf{y}_s, \omega) = [h(x_m, y_n|x_s^{(1)}, y_s^{(1)}, \omega) \; h(x_m, y_n|x_s^{(2)}, y_s^{(2)}, \omega) \cdots h(x_m, y_n|x_s^{(L)}, y_s^{(L)}, \omega)], \quad (8)$$

where $\mathbf{x}_s=[x_s^{(1)}x_s^{(2)}\cdots x_s^{(L)}]$ and $\mathbf{y}_s=[y_s^{(1)}y_s^{(2)}\cdots y_s^{(L)}]$. If we denote the input signals into the control sources as $\mathbf{q}(\omega)=[q^{(1)}(\omega)q^{(2)}(\omega)\cdots q^{(L)}(\omega)]^T$, the pressure at the position (x_m, y_n) is

$$p(x_m, y_n, \omega) = \mathbf{h}(x_m, y_n|\mathbf{x}_s, \mathbf{y}_s, \omega)\mathbf{q}(\omega). \quad (9)$$

The spatial correlation matrix (which is explained in Sec. II) of the bright zone is expressed in discrete form

$$\mathbf{R}_b(\omega) = \frac{1}{B(H_t-H)} \times \sum_{m=(B_t-B)/2\Delta x+1}^{(B_t+B)/2\Delta x} \sum_{n=H/\Delta y+1}^{H_t/\Delta y} \mathbf{h}(x_m, y_n|\mathbf{x}_s, \mathbf{y}_s, \omega)^H \times \mathbf{h}(x_m, y_n|\mathbf{x}_s, \mathbf{y}_s, \omega)\Delta x\Delta y. \quad (10)$$

The spatial correlation matrix of the dark zone is

$$\mathbf{R}_d(\omega) = \frac{1}{(B_t-B)Ht} \times \sum_{m=1}^{(B_t-B)/2\Delta x} \sum_{n=1}^{H_t/\Delta y} \mathbf{h}(x_m, y_n|\mathbf{x}_s, \mathbf{y}_s, \omega)^H \times \mathbf{h}(x_m, y_n|\mathbf{x}_s, \mathbf{y}_s, \omega)\Delta x\Delta y + \frac{1}{(B_t-B)H_t} \sum_{m=(B_t+B)/2\Delta x+1}^{(B_t/\Delta x)+1} \times \sum_{n=1}^{H_t/\Delta y} \mathbf{h}(x_m, y_n|\mathbf{x}_s, \mathbf{y}_s, \omega)^H \mathbf{h}(x_m, y_n|\mathbf{x}_s, \mathbf{y}_s, \omega)\Delta x\Delta y. \quad (11)$$

As shown in Sec. II, the problem to maximize the acoustic contrast is reduced to an eigenvalue problem as

$$\mathbf{R}_d(\omega)^{-1}\mathbf{R}_b(\omega)\mathbf{q}(\omega) = \beta(\omega)\mathbf{q}(\omega). \quad (12)$$

The contrast control solution $\mathbf{q}_c(\omega)$ is obtained as the eigenvector that corresponds to the maximum eigenvalue.

B. Constructing a multi-channel filter

The solution set which is obtained at each frequency is composed of complex numbers correspondent to channels. The solution sets at all frequencies of interest construct a multi-channel filter. A monosignal is filtered by the multi-channel filter and then the resultant signals are fed into the channels, respectively. The filtering process can be taken in time domain or frequency domain.

It is noteworthy that the frequency range has to be determined by the linear region of loudspeakers and the frequency interval has to be determined by considering required length of the impulse response of the solutions. In other words, in order to treat reverberations, time length is re-

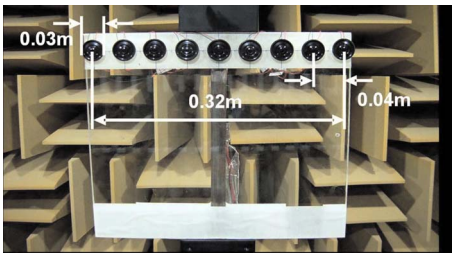


FIG. 5. (Color online) The loudspeaker array; the diameter of a loudspeaker is 0.03 m, the spacing between loudspeakers is 0.04 m, the number of loudspeakers is 9, and the aperture size is 0.32 m.

quired to be long, and then the frequency interval, which is the inverse of time length of the impulse response, has to be short.

V. THE REALIZATION OF A SYSTEM FOR 17 in. DISPLAY

In order to make sure that the method can be practically realizable, we need a well designed experiment. As a specific example of the personal audio system, we considered the 17 in. monitor display which is used in a laptop computer, desktop computer, or TV. We used loudspeakers whose diameter is 0.03 m as control sources. The experiment was done in an anechoic chamber $[(3.6^W \times 3.6^L \times 2.4^H \text{ m}^3)$; cut-off frequency: 100 Hz] to satisfy free-field assumption.

A. Experimental setup

Figure 5 shows the loudspeaker array that is used in the experiment. Considering the scale of 17 in. monitor display (the width is 0.32 m and height is 0.27 m), we regarded the aperture of control source array L_a to be 0.32 m. We used nine control sources ($L=9$) with equal spacing, 0.04 m, then the aperture size will be L_a, 0.32 m. Then the positions of control sources are

$$(x_s^{(l)}, y_s^{(l)}) = (-0.16 + 0.04(l-1), 0) \quad \text{(m)}, \tag{13}$$

$$l = 1, 2, \ldots, 9.$$

Figure 6 shows frequency response function $[\mathbf{h}(x_m, y_n | \mathbf{x}_s, \mathbf{y}_s, \omega)$ of Eq. (8)] and coherence between the input voltage signal into a loudspeaker unit and the output pressure (Pa) signal which is measured at a point 1 m ahead. Based on the figure, we selected the available frequency range to be 800 Hz–5 kHz because the coherence function is unity in the range; it means that the loudspeaker has linear response with respect to the input signal. The frequency range is discretized by $\Delta f = 10$ Hz because we determine the time length of impulse response to be 0.1 s. The time length is enough because we assume the free-field condition and we do not consider reflections. The number of the discretized frequencies N is 521.

We also assumed that B is 0.4 m. The H_t and H can be 0.6 and 0.2 m, respectively, to include the user's head, which is fairly large to allow the head movement. It is noteworthy that H has to be bigger than 0.07 m to satisfy $kH > 1$ for all frequency range that we have an interest in, 800 Hz–5 kHz. The width of total control zone B_t is decided to be 0.8 m.

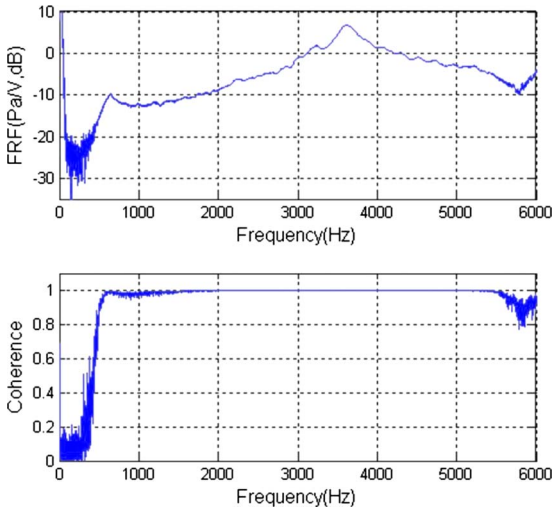


FIG. 6. (Color online) Frequency response function and coherence of a loudspeaker unit; the available frequency range is about 800 Hz–5 kHz.

Then the width of each dark zone (on the left and the right side of the bright zone) is 0.2 m, which includes the head of each person who can be right beside the user. The widths of the dark zones do not have to be big to include the head movement because the sound pressure in the outer region of the dark zone is as low as in the dark zone.

B. Measuring transfer functions and obtaining solutions

The measurement spacings Δx and Δy are decided to be 0.02 m, which is about 1/3 of wavelength of the highest frequency of interest, 5 kHz. Then the measurement position is

$$(x_m, y_n) = (-0.39 + 0.02(m-1), 0.01 + 0.02(n-1)) \quad \text{(m)}, \tag{14}$$

where $m = 1, 2, \ldots, 40$ and $n = 1, 2, \ldots, 30$.

We used 20 array microphones, which are spaced by 0.04 m to avoid scattering effect that is possible when the microphones are too close. Signal generating part is composed of PC, multi-channel sound card, multi-channel audio amp, and the loudspeaker array. Measuring part is composed of array microphones, analog-to-digital converter, and PC. In total, 60 steps (2 steps in x-axis and 30 steps in y-axis) of measurements were performed.

At each frequency, we have the transfer function $\mathbf{h}(x_m, y_n | \mathbf{x}_s, \mathbf{y}_s, \omega)$ [Eq. (8), the size 1×9] at all measurement points [Eq. (7), 40×30 points]. For example, Fig. 7 represents absolute value of transfer function by the first loudspeaker. From this transfer function, we can obtain the solution as explained in Sec. II. We obtain the spatial correlation matrices $\mathbf{R}_b(\omega)$ and $\mathbf{R}_d(\omega)$ [Eqs. (10) and (11); their size is 9×9] and the contrast control solution $\mathbf{q}_c(\omega)$ by solving the eigenvalue problem.

C. Experimental results

Then the resultant pressure at each point $p(x_m, y_n, \omega)$ is obtained by multiplication of the transfer function

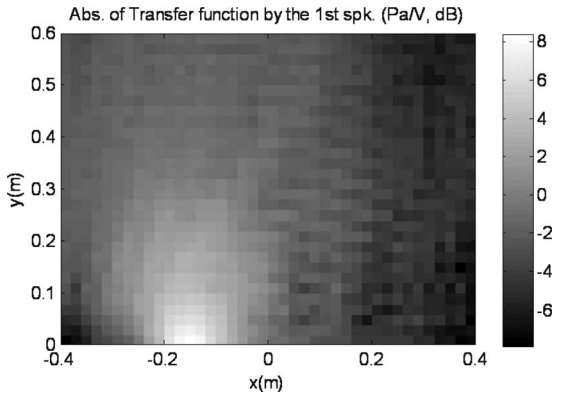


FIG. 7. Absolute values of transfer function between input voltage (V) and sound pressure (Pa) on the sound field of interest by the first loudspeaker, which is located at (−0.16 m, 0 m).

$\mathbf{h}(x_m, y_n | \mathbf{x}_s, \mathbf{y}_s, \omega)$ and the solution $\mathbf{q}_c(\omega)$ [Eq. (9)]. It is assumed to be same as actual pressure because this system is linear.

In order to evaluate this solution, we compare it with two other cases: Uncontrolled case and time reversal array case. The solution in the uncontrolled case $\mathbf{q}_{EI}(\omega)$ is determined by

$$\mathbf{q}_{EI}(\omega) = \sqrt{\frac{\sum_{l=1}^{L} |q_c^{(l)}(\omega)|^2}{L}} [1 \quad 1 \cdots 1]^T, \tag{15}$$

which is normalized for its power to be that of $\mathbf{q}_c(\omega)$. The solution has equal inputs into all channels. On the other hand, the solution in the time reversal array case $\mathbf{q}_{\mathrm{TRA}}(\omega)$ is determined by

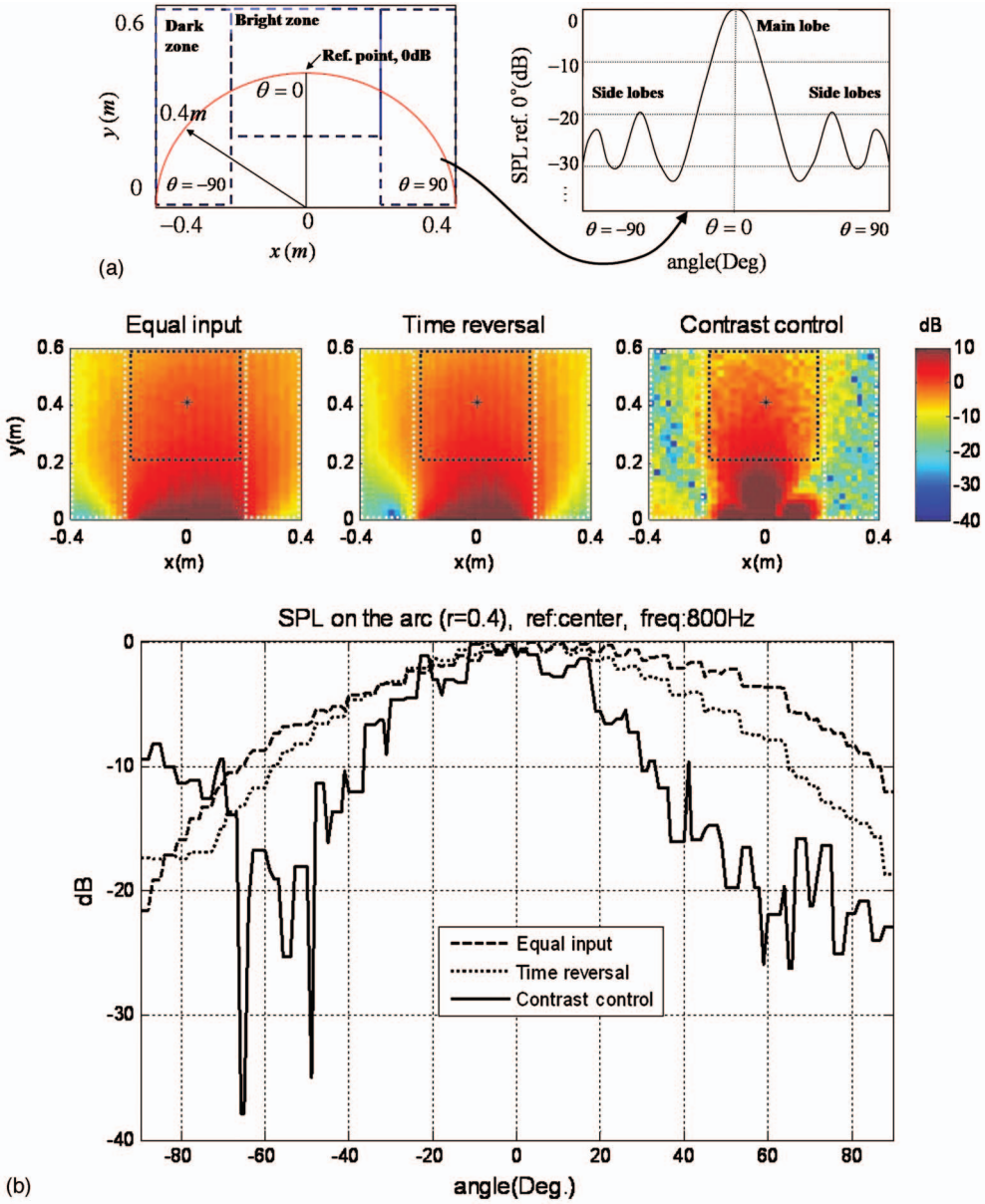


FIG. 8. The experiment results at the selected frequencies: 800 Hz and 5 kHz. (a) Sound field and its angular characteristics on the arc (0.4 m radius): The magnitude of sound pressure is expressed in decibel scale when the reference value is the pressure at the center (0.0 1 m, 0.4 m). The angle is 90° on the $+x$ axis and −90° on the $-x$ axis. The angle increases clockwise. (b) The results at 800 Hz; the top three figures show sound fields by equal input, time reversal, and contrast control, respectively. The mark " *" on the figures represents the position of the reference point. The center rectangle indicates the bright zone and the other rectangles at the sides indicate the dark zone. The bottom figure shows the angular characteristics in the three cases.

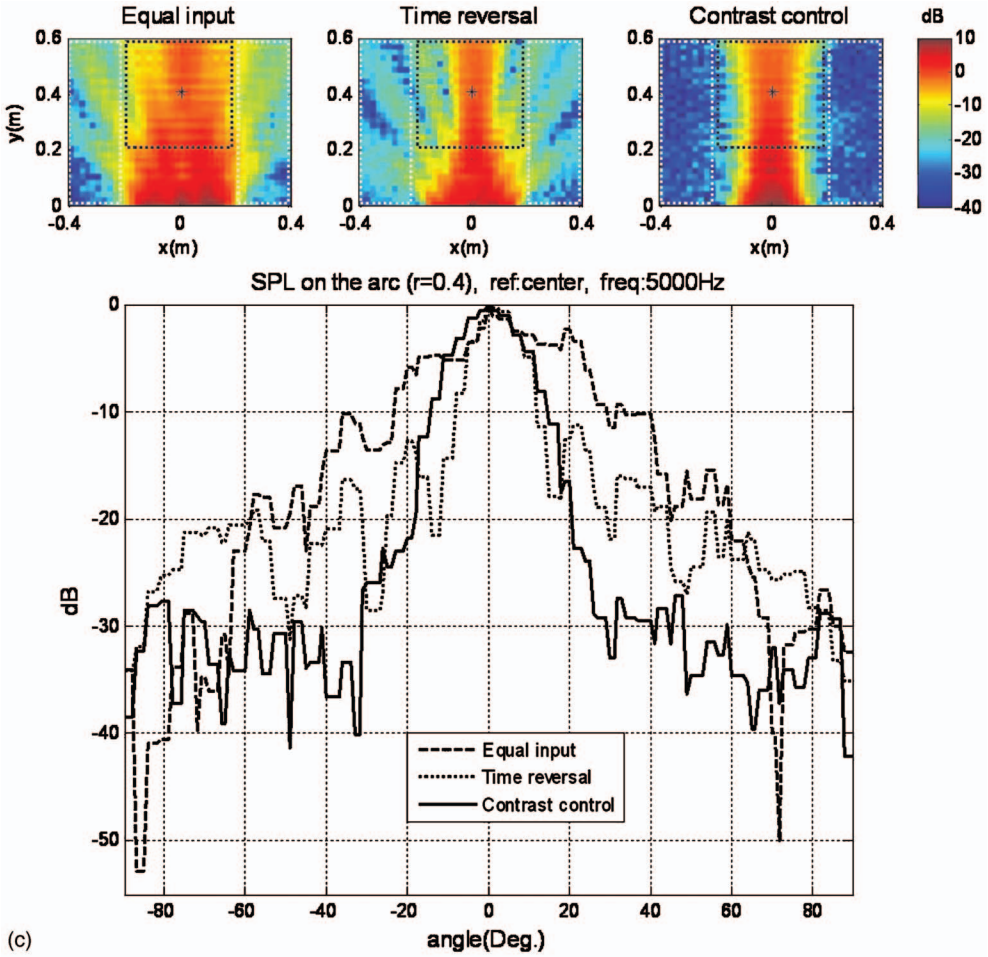


FIG. 8. (c) The results at 5 kHz; the structures of the figures are the same as what is described in the above Fig. 8(b).

$$\mathbf{q}_{TRA}(\omega) = \sqrt{\frac{\sum_{l=1}^{L} |q_c^{(l)}(\omega)|^2}{\sum_{l=1}^{L} |h(x_0, y_0 | x_s^{(l)}, y_s^{(l)}, \omega)|^2}} \times \mathbf{h}(x_0, y_0 | \mathbf{x}_s, \mathbf{y}_s, \omega)^H, \quad (16)$$

which is normalized so that its power is equal to that of $\mathbf{q}_c(\omega)$ and (x_0, y_0) is the location of the center of the user, (0.01 m, 0.4 m). [In fact, (0.01 m, 0.4 m) is considered as the center point because the center (0 m, 0.4 m) does not belong to the measurement points.]

Figure 8 shows the results of the experiments at the two selected frequencies: 800 Hz and 5 kHz. As illustrated in the left figure of Fig. 8(a), the top three figures of Figs. 8(b) and (c) show sound fields by equal input (uncontrolled case), time reversal, and contrast control, respectively. The sound pressure level is expressed by color. The center rectangle indicates the bright zone and the other rectangles at the sides indicate the dark zone. The magnitude of sound pressure is expressed in decibel scale when the reference value is the pressure at the center. The bottom figure shows the angular characteristics in the three cases. As shown in the right figure of Fig. 8(a), the sound pressure on the arc whose radius is 0.4 m is plotted with respect to the angle. (In fact, measurement points that are the closest ones to the arc are chosen.) The angle is 90° on the $+x$ axis and $-90°$ on the $-x$ axis, and increases clockwise. Sound pressure levels in the three cases are represented, respectively.

At all frequencies, the upper figures of Figs. 8(b) and (c) show that the sound pressure levels of the dark zones in the contrast control are lower than those in the equal input cases (uncontrolled cases) and time reversal array cases. The lower figures of Figs. 8(b) and (c) show that most levels of side lobes in contrast control are lower than those in the equal input cases and time reversal array cases. In the contrast control, except 800 Hz, the levels of most side lobes are lower than -20 dB.

Figure 9 shows the acoustic contrast in the three cases in the entire frequency region of interest, 800 Hz–5 kHz. In the contrast control case, the contrasts are more than 19 dB and higher than those in the other cases. Especially as frequency increases, the contrast in the contrast control case increases gradually while the contrasts in the other cases do not; as frequency goes high, the difference of the contrasts in the contrast control and the others has tendency to increase.

Due to the limited frequency range, sound quality for this system is inadequate to play music sounds, but speech sounds can be recognized. The frequency range of this system depends on that of loudspeakers, which has to do with the size of loudspeaker units. In other words, the size of

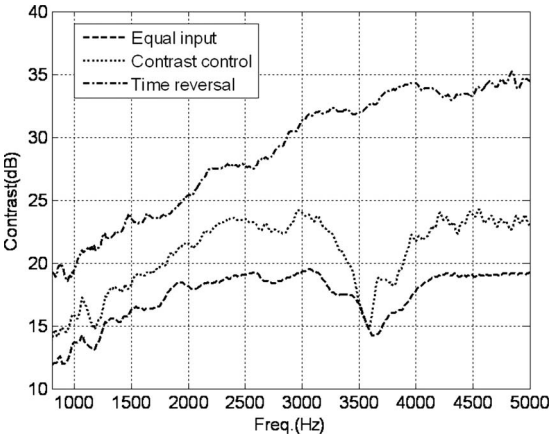


FIG. 9. The contrast in the three cases: Equal input (uncontrolled case), acoustic contrast control, and time reversal array.

loudspeakers, which are used in this system, is too small to generate the low frequency sound below 800 Hz. Therefore, in order to improve the sound quality, larger loudspeakers have to be used, and then the size of this system will also be larger.

VI. CONCLUSION

Acoustic contrast control was applied to develop a sound focused personal audio system. A line array of control sources was considered. As a specific example of the system, we considered a line array that can be attached on 17 in. monitor display. The interest frequency range was from 800 Hz to 5 kHz. Experimental results show that the cases of acoustic contrast control provide lower levels of side lobes and higher contrasts than the equal input case and time reversal array at all frequencies. In the contrast control, the levels of most side lobes are lower than −20 dB at the selected frequencies and the contrasts are more than 19 dB at all frequencies. In addition, as frequency goes high, the difference of the contrasts in the contrast control and the others has tendency to increase.

ACKNOWLEDGMENT

This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the National Research Laboratory Program funded by the Ministry of Science and Technology (Grant No. R0A-2005-000-10112-0).

[1]J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," J. Acoust. Soc. Am. **111**, 1695–1700 (2002).
[2]P. J. Westervelt, "Parametric acoustic array," J. Acoust. Soc. Am. **35**, 535–537 (1963).
[3]M. Fink, "Time reversal of ultrasonic fields. I. Basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–566 (1992).
[4]M. Tanter, J.-L. Thomas, and M. Fink, "Time reversal and the inverse filter," J. Acoust. Soc. Am. **108**, 223–234 (2000).
[5]S. Uṇṇikrishṇa Pillai, *Array Signal Processing* (Springer-Verlag, New York, 1989), pp. 15–20.

Beamforming with microphone arrays for directional sources

Christian Bouchard[a)]
School of Information Technology and Engineering, University of Ottawa, 800 King Edward, Ottawa, Ontario K1N 6N5, Canada, and Institute for Microstructural Sciences, National Research Council of Canada, Building M-36, Ottawa, Ontario K1A 0R6, Canada

David I. Havelock
Institute for Microstructural Sciences, National Research Council of Canada, Building M-36, Ottawa, Ontario K1A 0R6, Canada

Martin Bouchard
School of Information Technology and Engineering, University of Ottawa, 800 King Edward, Ottawa, Ontario K1N 6N5, Canada

(Received 4 September 2008; revised 29 January 2009; accepted 4 February 2009)

Beamforming is done with an array of sensors to achieve a directional or spatially-specific response by using a model of the arriving wavefront. Real acoustic sources may deviate from the conventional plane wave or monopole model, causing decreased array gain or a total breakdown of beamforming. An alternative to beamforming with the conventional source model is presented which avoids this by using a more general source model. The proposed method defines a set of "sub-beamformers," each designed to respond to a different spatial mode of the source. The outputs of the individual sub-beamformers are combined in a weighted sum to give an overall output of better quality than that of a conventional (monopole) beamformer. It is shown that with appropriate weighting, the optimum array gain can be achieved. A simple method is demonstrated to estimate the weighted sum, based on the observed data. The variance and bias of the estimate in the presence of noise are evaluated. Simulation and experimentally measured results are shown for a simple directive source. In the experiment, the proposed method provides an array gain of about 11 dB while beamforming using a point source model achieves only −4 dB.
 [DOI: 10.1121/1.3089221]

PACS number(s): 43.60.Fg, 43.60.Dh [EJS] Pages: 2098–2104

I. INTRODUCTION

Beamforming is done with an array of sensors to achieve a directional or spatially-specific response. It relies on a model of the wavefront (source model) arriving at the array to calculate the time delay, or frequency domain phase shift, that must be applied to the signal of each sensor so that they may be summed coherently. In what is often called near-field beamforming,[1] a point source (monopole) model is typically used. If the source is distant from the array, wavefront curvature is neglected (far-field assumption).

We are concerned here with sources that do not radiate as point sources and for which a different source model must be used to compute the beamforming weights.[2,3] The source may have an elaborate directivity pattern caused by a complex sound production mechanism, as for a musical instrument,[4,5] or it may consist of spatially distributed radiating elements with arbitrary but fixed phase relationships. These deviations from the point source model are most noticeable when the source subtends a large angle about the array. They may result in a decrease in array gain or a total breakdown of beamforming. To avoid this, a more realistic source model is required than is typically seen in the literature or done in practice.

Beamforming based directly on a specific source model is referred to here as *conventional beamforming*. With this method, the proper source model would have the same directivity and orientation as the actual source; however, source directivity and orientation are unknown in most cases. This paper introduces an alternative approach to conventional beamforming. The general principle is to define a set of "sub-beamformers," each beamforming on a different spatial mode of the source. The outputs of the individual sub-beamformers are combined in a weighted sum to give an overall output of better quality than that of a conventional beamformer using the monopole model. This paper shows that, with the appropriate weighting, the sum of the sub-beamformers gives the same optimum array gain as conventional beamforming with the exact source model. A method to estimate the sub-beamformer weights based on the observed data is demonstrated. The bias and variance of the estimate in the presence of noise are evaluated.

The proposed method has the following benefits when compared with beamforming using a point source model:

- it can overcome deteriorated beamformer response that occurs when a null of the source directivity falls on the receiving array,[3]
- in a stationary environment, once the sub-beamformer weights are evaluated, the computational complexity can be made the same by combining the sub-beamformers,

[a)]Electronic mail: christian.bouchard@nrc-cnrc.gc.ca

- it has similar performance when beamforming on point sources, and
- it works with the same array geometry and hardware.

Beamforming can be used for spatial differentiation, tracking, or high quality sound pickup at a distance from the source. The proposed beamforming method is developed to address the latter objective. Unlike near-field acoustical holography[6] (NAH) or other techniques,[7,8] the method does not address reconstruction of the acoustic pressure field near the source or evaluation of the surface velocity of the source. Furthermore, unlike NAH, the near-field criterion for beamforming is wavefront curvature rather than the presence of evanescent components.

Here, we introduce the method and demonstrate through simulation and experiment its advantages over conventional beamforming with a point source. Detailed quantification of performance is beyond the scope of this paper and so this important analysis is left to future work. In general, however, the performance of the proposed method depends on the source directivity and position, on the array geometry, and the environment in which the array is operating. Reverberation and interfering sources can also be expected to have a strong impact on performance, particularly in regard to estimation of the sub-beamformer weights.

The principles of beamforming and array noise gain optimization are reviewed in Sec. II, and the proposed method is introduced in Sec. III. Section IV presents an estimator of the weight to give each sub-beamformer in the linear sum of the proposed method and provides the bias and variance of this estimator for a known noise field. In Sec. V it is shown that if the source directivity pattern and source location with respect to the array do not change with time then it is possible to obtain the overall beamformer output without computing the output of each sub-beamformer. In this case, the method is computationally equivalent to beamforming directly with a source model. Section VI compares simulations and measurements for a simple directive source when beamforming with both the proposed method and using the point source model. Section VII compares the proposed method with the phase-mode spherical array[9–11] and shows that the methods are quite different even though they both use spherical harmonic expansions. Concluding remarks are in Sec. VIII.

II. ARRAY BEAMFORMING AND OPTIMIZATION

The complex exponential $e^{i2\pi ft}$ is used here for the time dependence. Where no ambiguity exists, the continuous and discrete cases are distinguished by the use of t or n as the time parameter and the use of f or k as the frequency parameter, respectively.

The forward Fourier transform is applied to the time-domain signal $p_l(t)$ at the lth microphone to obtain $p_l(f)$ in the frequency domain. In the discrete (time sampled) case the normalized discrete Fourier transform is used to obtain $p_l(k)$ (normalization by $\sqrt{N}$, where N is the number of time samples used in the transform).

In the frequency domain, the beamformed output of an array of L microphones is

$$y(f) = \sum_{l=1}^{L} w_l^*(f) p_l(f), \tag{1}$$

where $w_l(f)$ is the complex valued beamforming weight for the lth microphone. The asterisk denotes complex conjugation.

The output of microphone l, located at the point (r_l, θ_l, ϕ_l), is composed of the signal of interest, s, and additive noise e_l,

$$p_l(f) = s(f, r_l, \theta_l, \phi_l) + e_l(f). \tag{2}$$

The noise is assumed to be uncorrelated with the signal. Using vector notation, the beamforming weights are assembled in a vector $\mathbf{w}(f)$ and the microphone outputs in a vector $\mathbf{p}(f) = \mathbf{s}(f) + \mathbf{e}(f)$. The vector $\mathbf{s}(f)$ is called the *signal model*. The array output is

$$y(f) = \mathbf{w}^H(f)\mathbf{p}(f). \tag{3}$$

The superscript H indicates the Hermitian conjugate transpose. The beamforming weight vector may be optimized in regard to some criteria, such as the signal-to-noise ratio (SNR) or directional response pattern.

The *gain* of the microphone array is defined as the ratio of the SNR at the output of the beamforming array to the SNR of a single reference microphone. For development of the theory, the reference SNR is defined here, as it is in Ref. 12, to be the ratio of the average signal power spectral density over the L microphones of the array, $\sigma_{\mathbf{s}}^2(f) = \mathbf{s}^H(f)\mathbf{s}(f)/L$, to the average noise power spectral density over the array, $\sigma_{\mathbf{e}}^2(f) = E\{\mathbf{e}^H(f)\mathbf{e}(f)\}/L$. The resulting expression for the gain of the array at a specific frequency is

$$G(f) = \frac{\mathbf{w}^H(f)\mathbf{R}_{\mathbf{ss}}(f)\mathbf{w}(f)}{\mathbf{w}^H(f)\mathbf{R}_{\mathbf{ee}}(f)\mathbf{w}(f)}, \tag{4}$$

where $\mathbf{R}_{\mathbf{ss}}(f) = \mathbf{s}(f)\mathbf{s}^H(f)/\sigma_{\mathbf{s}}^2(f)$ is called the normalized signal cross-power spectral density matrix, and $\mathbf{R}_{\mathbf{ee}}(f) = E\{\mathbf{e}(f)\mathbf{e}^H(f)\}/\sigma_{\mathbf{e}}^2(f)$ is the normalized noise cross-power spectral density matrix. Provided $\mathbf{R}_{\mathbf{ee}}(f)$ is not singular, the array gain is maximized with respect to SNR with the weight vector[12]

$$\mathbf{w}_{\text{opt}}(f) = \alpha(f)\mathbf{R}_{\mathbf{ee}}^{-1}(f)\mathbf{s}(f). \tag{5}$$

The scaling factor $\alpha(f)$ is an arbitrary complex constant and it is set to unity in the remainder of this paper. When the noise at the microphone outputs is additive white-noise uncorrelated from sensor to sensor, such as in spatially uncorrelated white-noise, $\mathbf{R}_{\mathbf{ee}}(f)$ is equal to the identity matrix $\mathbf{I}_L$ and the optimum weight vector becomes equal to the signal model itself

$$\mathbf{w}_{\text{opt}}(f) = \mathbf{s}(f) \quad \text{when} \quad \mathbf{R}_{\mathbf{ee}} = \mathbf{I}_L. \tag{6}$$

In this case, the array gain is called *white-noise gain*.

For the point source model, the optimum weight at the lth microphone to maximize the white-noise gain is

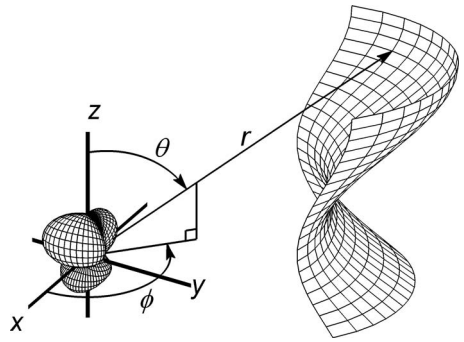


FIG. 1. Arbitrary discrete array and source with an arbitrary directivity (magnitude shown, phase not shown).

$$w_{\mathrm{opt},l}(f) = \frac{1}{r_l} e^{-i2\pi f r_l/c}, \tag{7}$$

where r_l is the distance from microphone l to the source and c is the speed of sound. In the time domain, this is a delay-and-sum beamformer

$$y(t) = \sum_{l=1}^{L} \frac{1}{r_l} p_l(t + t_l), \tag{8}$$

where $t_l = r_l/c$ is the propagation time to reach the lth microphone.

III. DEVELOPMENT OF THE PROPOSED METHOD

Figure 1 illustrates schematically the directivity magnitude of an arbitrary source, at some frequency f, and a discrete sensor array with some arbitrary geometry. (The phase of the directivity is not shown.) For notational convenience, and without loss of generality, the source is located at the center of the spherical coordinate system.

Using a decomposition[6] into spherical harmonic modes (n,m) at frequency f, the outgoing wave from a source generates a signal at a point (r,θ,ϕ),

$$s(f,r,\theta,\phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} s_{nm}(f) h_n^{(2)}(\kappa r) Y_n^m(\theta,\phi), \tag{9}$$

where $h_n^{(2)}(\kappa r)$ is the Hankel function of the second kind of order n, $\kappa = 2\pi f/c$ is the wave number, and $Y_n^m(\theta,\phi)$ is the spherical harmonic function of order n and degree m. The factor $s_{nm}(f)$ in the summation is a complex number equal to the contribution of the spherical harmonic mode (n,m) from the source at frequency f; it is referred to here simply as the *mode strength*.

The directivity of the source, and so the mode strengths s_{nm}, is often unknown. This section demonstrates that if these mode strengths can be determined then a properly weighted linear sum of specially defined beamformers (called *sub-beamformers*) provides a response equivalent to that of conventional beamforming with the exact source model. In this case, the optimum array gain is achieved.

For a discrete and time sampled array, the signal in the kth frequency bin of the lth sensor is

$$s_l(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} s_{nm}(k) h_n^{(2)}(\kappa_k r_l) Y_n^m(\theta_l,\phi_l), \tag{10}$$

where κ_k is the wave number at the center of frequency bin k. A *signal model vector* $\mathbf{s}(k)$ is formed from the signal component of all microphones,

$$\mathbf{s}(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} s_{nm}(k) \mathbf{Y}_n^m(k), \tag{11}$$

where, for notational convenience, we also define the *mode vector*

$$\mathbf{Y}_n^m(k) = [h_n^{(2)}(\kappa_k r_1) Y_n^m(\theta_1,\phi_1) \cdots h_n^{(2)}(\kappa_k r_L) Y_n^m(\theta_L,\phi_L)]^T. \tag{12}$$

The reader is cautioned that the letter Y is used both for the mode vector and for the spherical harmonic function, albeit in different font properties, and that the mode vector $\mathbf{Y}_n^m(k)$ contains both the spherical harmonic and the Hankel function of the second kind.

The weight vector to maximize the array gain given by Eq. (5) is

$$\mathbf{w}(k) = \mathbf{R}_{\mathbf{ee}}^{-1}(k) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} s_{nm}(k) \mathbf{Y}_n^m(k), \tag{13}$$

or, rearranging,

$$\mathbf{w}(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} s_{nm}(k) \mathbf{R}_{\mathbf{ee}}^{-1}(k) \mathbf{Y}_n^m(k). \tag{14}$$

Inserting this weight vector into Eq. (3), the optimized array beamformer output is

$$y(k) = \left\{ \sum_{n=0}^{\infty} \sum_{m=-n}^{n} s_{nm}(k) \mathbf{R}_{\mathbf{ee}}^{-1}(k) \mathbf{Y}_n^m(k) \right\}^H \mathbf{p}(k) \tag{15a}$$

$$= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \underbrace{s_{nm}^*(k)} \underbrace{\{\mathbf{Y}_n^{mH}(k) \mathbf{R}_{\mathbf{ee}}^{-1}(k) \mathbf{p}(k)\}} \tag{15b}$$

$$= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \{s_{nm}(k)\}^* \{y_{nm}(k)\}. \tag{15c}$$

Each factor $y_{nm}(k)$ in the curly bracket in Eq. (15c) has the form of a beamformer [viz., Eq. (3)] with beamforming weight vector $\mathbf{w}_{nm}(k) = \mathbf{R}_{\mathbf{ee}}^{-1}(k) \mathbf{Y}_n^m(k)$. Considering the form of the gain-optimized array weight vector given in Eq. (5), the beamformer $y_{nm}(k)$ is called a *gain-optimized sub-beamformer* of order n and degree m.

Since the gain-optimized sub-beamformers are readily computed from the signals at the microphone array, then the linear sum of sub-beamformers in Eq. (15c) can be used to perform beamforming with optimal gain if the weights $s_{nm}(k)$ can be found somehow. Figure 2 shows a schematic of the proposed method with the array data processed in parallel through a set of sub-beamformers and the outputs $y_{nm}(k)$ multiplied by $s_{nm}^*(k)$ and summed.

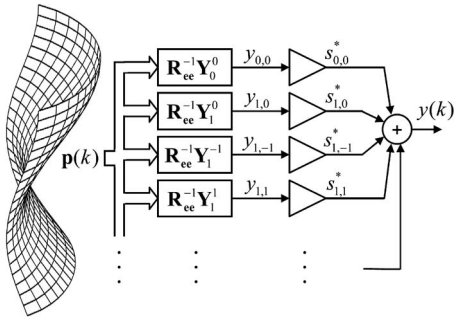


FIG. 2. Proposed method schematic.

It is noteworthy that if the source can be decomposed using some other set of modes $\mathbf{M}_m(k)$ as

$$\mathbf{s}(k) = \sum_{m=0}^{M} s_m(k)\mathbf{M}_m(k), \tag{16}$$

then the method can be applied using

$$y_m(k) = \mathbf{M}_m^H(k)\mathbf{R}_{\mathbf{ee}}^{-1}(k)\mathbf{p}(k). \tag{17}$$

IV. ESTIMATING THE MODE STRENGTHS

Since the strength of each mode of the source is generally unknown, it is necessary to estimate $s_{nm}(k)$ in Eq. (15c). If the source has only a single mode, (n,m), with strength $s_{nm}(k)$, then the sub-beamformer output for that mode would be

$$y_{nm}(k) = \mathbf{Y}_n^{mH}(k)\mathbf{R}_{\mathbf{ee}}^{-1}(k)\{s_{nm}(k)\mathbf{Y}_n^m(k) + \mathbf{e}(k)\} \tag{18}$$

Based on this observation, and defining the coefficient

$$C_{n'm'}^{nm}(k) = \mathbf{Y}_n^{mH}(k)\mathbf{R}_{\mathbf{ee}}^{-1}(k)\mathbf{Y}_{n'}^{m'}(k), \tag{19}$$

we define a possible estimator for the mode strength as

$$d_{nm}(k) = \frac{1}{C_{nm}^{nm}(k)} y_{nm}(k). \tag{20}$$

If the source has other modes, then this estimator of the (n,m) mode strength will be biased. The bias is a weighted sum of the other mode strengths

$$E\{d_{nm}(k)\} - s_{nm}(k) = \frac{1}{C_{nm}^{nm}(k)} \sum_{\substack{n'=0 \\ (n',m')\neq(n,m)}}^{\infty} \sum_{m'=-n'}^{n'} C_{n'm'}^{nm}(k) s_{n'm'}(k). \tag{21}$$

Notice that the bias is independent of the average noise power over the array.

The variance of the estimator is proportional to the average noise power over the array

$$\mathrm{Var}\{d_{nm}(k)\} = \frac{\sigma_{\mathbf{e}}^2(k)}{\mathbf{Y}_n^{mH}(k)\mathbf{R}_{\mathbf{ee}}^{-1}(k)\mathbf{Y}_n^m(k)} = \frac{\sigma_{\mathbf{e}}^2(k)}{C_{nm}^{nm}(k)}. \tag{22}$$

We have assumed that the noise field, and hence $\mathbf{R}_{\mathbf{ee}}(k)$, is known; in practice, however, $\mathbf{R}_{\mathbf{ee}}(k)$ would have to be estimated. In this case, the bias and variance would need to take into account the characteristics of the estimator of $\mathbf{R}_{\mathbf{ee}}(k)$.

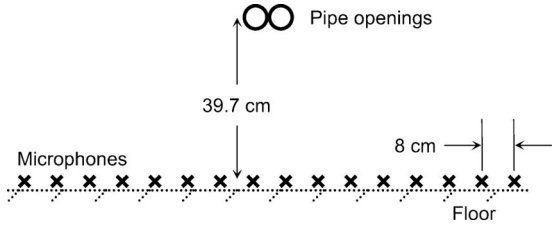


FIG. 3. Experiment setup.

Reverberation can be expected to degrade the mode strength estimation but this is not addressed here.

V. COMPUTATIONAL COMPLEXITY

The beamforming weight vector $\mathbf{w}_{nm}(k)$ and the source mode strength estimate $d_{nm}(k)$ must be estimated for each sub-beamformer. The number of such parameters to be estimated increases with the number N_{order} of spherical harmonic mode orders being considered. In practice, N_{order} should be determined by the expected complexity of the source. If the source directivity pattern and source location with respect to the array do not change with time, once the source mode strengths are estimated it is possible to reduce the amount of computation necessary to obtain the overall beamformer. This is done by computing an overall beamforming weight vector using Eq. (14) with the estimate $d_{nm}(k)$ replacing $s_{nm}(k)$ which is then used to beamform, without requiring the computation of the individual sub-beamformer outputs. The resultant beamformer output is

$$y(k) = \underbrace{\left\{\sum_{n=0}^{N_{\mathrm{order}}} \sum_{m=-n}^{n} d_{nm}(k)\mathbf{R}_{\mathbf{ee}}^{-1}(k)\mathbf{Y}_n^m(k)\right\}^H}_{=\ \mathbf{w}^H(k)} \underbrace{\mathbf{p}(k)}_{\mathbf{p}(k)}. \tag{23}$$

VI. SIMULATION AND MEASUREMENT RESULTS

This section describes the simulation and experimental demonstration of the proposed method using a directional source and a microphone array. The setup of the demonstration is shown schematically in Fig. 3.

The linear array has 16 uniformly spaced 0.25-in. (6 mm) electret microphones with 8-cm spacing. The array is positioned on the floor with the microphone centers approximately 9 mm above the floor. The source is constructed with two 1-in. (26 mm) PVC (polyvinyl chloride) pipes, approximately 3 m long, each connected to a compression driver. The pipes are approximately parallel with their open ends in contact. The centers of the opening are $h=39.7$ cm directly above the microphone array and in line with the axis of the array.

A 1-kHz test signal is sent to each compression driver. The input signals to the drivers are always 180 deg out of phase with respect to each other and their amplitudes are independently adjustable. The acoustic signal measured at the microphone array includes the 1-kHz test tone, room re-

verberation, noise, and distortion. Identical drivers, pipes, and couplers were used to minimize differences between the two source channels so that a dipole source is approximated when the drive voltages are equal. Since reverberation (image sources) is inevitable in a non-anechoic environment, the microphones are positioned as close as possible to the floor (which is vinyl over wood frame) so that the dominant (nearest) image source is symmetrically positioned about the array axis and will not affect the apparent source directivity. To reduce the effects of other source images, various absorbing panels and plush furniture are positioned about the room, as distant as possible from the source and array. Due to the hard floor and ceiling, the reverberation time in the vicinity of the array is about 700 ms.

For analysis and simulation, a coordinate system is defined with the origin between the centers of the pipe openings, the xy-plane parallel to the floor, the x-axis parallel to the microphone array, and the z-axis passing through the array center, so that the array center is located at $(0,0,-h)$. The center of the pipe openings is located at $(\pm dx/2,0,0)$, where dx is the pipe diameter. The source due to each of the pipes is modeled as a point source at the effective position $(\pm dx/2,-dl,0)$, which is a small distance dl in front of the pipe end.[13] In the simulation computations, this small displacement of the source is modeled but reverberation and noise are ignored.

Digital signals are acquired from the microphone array using a 16-channel anolog-to-digital converter with 16-bit resolution and a sampling rate of 48 kHz. Processing is done offline using a 4096-point fast Fourier transform (FFT) with hamming windowing on each microphone signal to obtain the Fourier coefficient for the single FFT bin containing the 1-kHz test tone. The Fourier coefficients from the 16 microphones are assembled into the microphone signal vector $\mathbf{p}(k)$.

As an indication of the SNR, the FFT bin containing the 1-kHz tone for each of the 16 channels was observed to have a power spectral density at least 40 dB higher when the signal was present than when no signal was present.

In this demonstration of the proposed method, only sub-beamformers of order $N_{\text{order}}=4$ or less are considered. This number was chosen because it is more than adequate for a dipole and provides low computational complexity. Typically, some knowledge of the source characteristics would be used to guide the selection of a value for N_{order}. The contribution to each mode $d_{nm}(k)$ is estimated using Eq. (20), and the overall beamforming weight vector is then obtained as

$$\mathbf{w}(k)=\sum_{n=0}^{4}\sum_{m=-n}^{n} d_{nm}(k)\mathbf{Y}_n^m(k). \tag{24}$$

The optimized white-noise gain for the beamformer is computed as

$$G_{\text{out,dB}}(k)=10\log\left(\frac{\mathbf{w}^H(k)\hat{\mathbf{R}}_{\mathbf{ss}}(k)\mathbf{w}(k)}{\mathbf{w}^H(k)\mathbf{w}(k)}\right). \tag{25}$$

This differs slightly from the white-noise gain in Eq. (4) [with $\mathbf{R}_{\mathbf{ee}}(k)$ equal to the identity matrix] because in our measurements the matrix $\mathbf{R}_{\mathbf{ss}}(k)$ is unknown and is estimated from a single microphone signal vector $\mathbf{p}(k)$, which includes some noise. It is estimated as

$$\hat{\mathbf{R}}_{\mathbf{ss}}(k)=\frac{L}{\mathbf{p}^H(k)\mathbf{p}(k)}\cdot\mathbf{p}(k)\mathbf{p}^H(k). \tag{26}$$

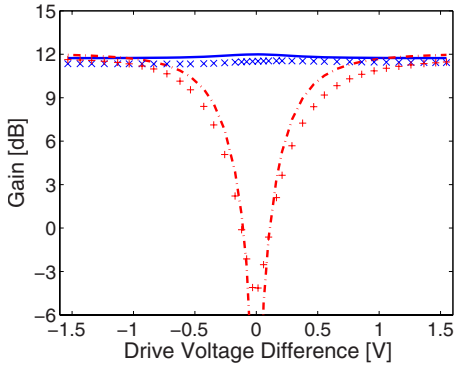


FIG. 4. (Color online) Array gain for varying drive voltage difference $V_{\text{driver\#2}}-V_{\text{driver\#1}}$. Driver No. 1 voltage is held at a fixed value of 1.534 V while driver No. 2 voltage is varied from 0 to 1.548 V, then driver No. 2 voltage is held at a fixed value of 1.548 V while driver No. 1 voltage is reduced to 0. Simulated (—) and measured (×) results for the proposed method, and simulated (·–) and measured (+) results for beamforming with a point source model.

If the drive voltage to either one of the drivers is reduced to zero, the overall source will change from a dipole to a monopole. Figure 4 shows the white-noise gain according to Eq. (25) for the measurement and simulation data, for the proposed method and beamforming with a point source model, when the drive voltages are changed. The pressure magnitude and phase at the end of each tube are not available so white-noise gain is plotted as a function of drive voltage difference. Note that for the proposed method, Eq. (24) is evaluated at each drive voltage whereas the point source model for conventional beamforming is kept constant for all drive voltages.

The optimum white-noise gain is constant; it depends only on the number of sensors and is independent of the source. From Eqs. (4) and (6), the optimum white-noise gain is $10\log(L)$, which is equal to 12.04 dB for an array of 16 sensors. As expected,[3,14] beamforming with a point source model deteriorates and does not perform properly when the drive voltages are equal and the source is a dipole. In which case, Fig. 4 shows a gain slightly below −4 dB for the measurement. On the other hand, for both the measurement and simulation, the proposed method maintains a gain of more than 11 dB.

At the source, only modes $(n,m)=(1,-1)$ and $(n,m)=(1,1)$ are necessary to generate a dipole with the orientation described for the setup. Table I presents the white-noise gain of each sub-beamformer in response to the dipole in the simulated case, computed using $\mathbf{w}(k)=\mathbf{Y}_n^m(k)$ in Eq. (25). It shows that numerous sub-beamformers have similar (but slightly lower) white-noise gain than the two required modes. This is because the linear array has a limited view of the dipole and so the source looks somewhat similar to other modes. (This would not be the case for a dense array completely surrounding the source.)

TABLE I. White-noise gain (in decibels) for each sub-beamformer $y_{nm}(k)$ in response to the dipole in the simulated case.

n	0	1	2	3	4
m	White-noise gain (dB)				
−4					−13.34
−3				11.37	11.61
−2			−20.59	−19.57	−18.35
−1		12.04	11.90	10.83	6.73
0	−17.78	−17.90	−19.04	−22.46	−26.33
1		12.02	11.88	10.80	6.67
2			−10.72	−10.05	−9.38
3				11.34	11.57
4					−8.43

As the overall source approaches a monopole (extreme left or right of the figure), conventional point source beamforming achieves slightly more gain than the proposed method.

A beampattern plot is often used to illustrate the spatial selectivity of a beamforming array using the conventional point source model. For conventional far-field beamforming (plane wave source model) the beampattern is simply a function of arrival direction while for near-field beamforming (point source model) it also depends on the distance from the source to the receiving array. These beampatterns can be easily displayed as one- or two-dimensional plots but spatial selectivity is much more difficult to depict for the proposed method because it is also dependent on the source directivity.

To provide some indication of spatial selectivity for the proposed method, Fig. 5 illustrates how the response varies with the source position when the overall source approximates a dipole. It presents results for source locations at fixed height above the array and along the length of the array. (This is actually done by shifting the array.) The beamforming weight vector for the proposed method is computed using Eqs. (20) and (24), with the source over the center of the array. The beamforming weight vector is then held constant for all other positions. Similarly, the beamforming weight vector for the conventional beamformer is based on a point source over the array center and is held fixed.

There is a good match between measurements and simulation results in Fig. 5. The side lobe levels obtained with the proposed method and with the conventional beamformer using the point source model are not greatly different but the proposed method overcomes the failure to sense the dipole source at the proper position.

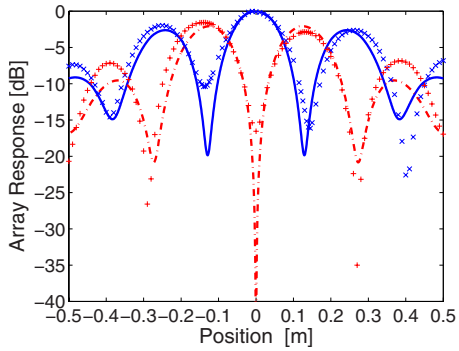


FIG. 5. (Color online) Array response along the x-axis for a dipole source. Simulated (—) and measured (×) results for the proposed method, and simulated (·–) and measured (+) results for beamforming with a point source model.

VII. COMPARISON OF THE PROPOSED METHOD WITH THE PHASE-MODE SPHERICAL ARRAY

The phase-mode spherical array[9–11] has some similarities in form to the proposed method. Both methods use spherical harmonic decomposition and perform a linear combination of eigenbeamformers or sub-beamformers (respectively). Beyond this, the two methods are very different.

The eigenbeamformers mathematically decompose the signal received at the array using spherical harmonics, whereas the proposed method decomposes the source itself. The eigenbeamformers use a spherical array to make the decomposition straightforward, whereas the proposed method applies to any array geometry. The phase-mode spherical array applies to known source models such as plane-waves or point sources whereas the proposed method addresses more general and unknown sources.

The number of modes used by the phase-mode spherical array is determined by the desired accuracy of approximation to beamforming directly with a point source model, and the optimization criterion, whereas the number of modes used with the proposed method is determined by the source complexity. For example, with a monopole source, the phase-mode spherical array requires an infinite number of modes to obtain the optimum white-noise gain, whereas the proposed method with one mode reduces identically to the conventional beamformer using the point source model.

The beamforming weights of each eigenbeamformer are fixed whereas the weight vector of each sub-beamformer depends on the source position and orientation with respect to the array, the array geometry, and the frequency. The weight vector also depends on the normalized noise cross-power spectral density matrix $\mathbf{R}_{\mathbf{ee}}(k)$ when maximizing the array gain.

The weights used in the output stage of the phase-mode spherical array depend on the source position, the array radius, the frequency, and the optimization criterion whereas

the weights to obtain the overall output of the proposed method depend only on the source mode strengths and optimization criterion.

The proposed method addresses a problem with significantly more degrees of freedom (source directivity and orientation) than beamforming with the conventional point source model and this is reflected in the need to estimate more parameters for the same order of harmonic decomposition.

VIII. CONCLUDING REMARKS

Simulation and experimentally measured results were shown for a simple directive source. In the experiment, when the source approached a dipole, the proposed method provided an array gain of about 11 dB while beamforming using a point source model achieved only -4 dB.

Optimization has been considered with respect to the noise field and is based on knowledge of the normalized noise cross-power spectral density matrix $\mathbf{R}_{\mathbf{ee}}(k)$. For clarity and simplicity, the proposed method has been introduced without considering possible techniques for estimating this matrix.

The mode vectors $\mathbf{Y}_n^m(k)$ are determined by the beamforming position; hence these vectors must be updated as the beamformer is steered to different positions. It has been assumed that the source position is known and position estimation is not addressed here.

A simple, intuitive estimator of the source mode strengths is proposed in Sec. IV, and the bias and variance of this estimate are analyzed. It may be possible to devise alternative estimators to reduce the bias or variance; however, the proposed estimator is computationally simple.

The proposed method is inherently adaptive in that an estimate of the unknown source mode strengths is obtained from the observed data. Reflections (reverberation) may affect these estimates and generally degrade beamforming performance. If the source directivity (and position) is known *a priori*, so that estimation is not required, then reflections behave as additional virtual sources that are reduced by the directivity of the proposed beamformer, just as for beamforming with the conventional source model.

The spherical harmonic orders included in the overall beamformer must be limited to a finite number N_{order}. This value should be estimated from the general complexity of the source. For the dipole example considered here, $N_{\text{order}}=4$ was found to be adequate.

The proposed method can be applied easily to situations of well known source directivity and noise. It constitutes a step toward beamforming in more general applications on unknown directional sources. Further work should address obtaining improved estimates of the mode strength, the number of modes required, and $\mathbf{R}_{\mathbf{ee}}(k)$.

ACKNOWLEDGMENTS

This work was supported in part by NSERC and NRC scholarships.

[1] J. G. Ryan and R. A. Goubran, "Application of near-field optimum microphone arrays to hands-free mobile telephony," IEEE Trans. Veh. Technol. **52**, 390–400 (2003).
[2] C. Bouchard, D. I. Havelock, and M. Bouchard, "Directional sources and beamforming," Proceedings of Meetings on Acoustics **5**, 055001 (2008).
[3] P. Jordan, J. A. Fitzpatrick, and J.-C. Valière, "Measurement of an aeroacoustic dipole using a linear microphone array," J. Acoust. Soc. Am. **111**, 1267–1273 (2002).
[4] C. Lheureux, "Simulation et mesure du rayonnement de différents instruments à vent à trous latéraux (Simulation and measurement of the radiation of lateral toneholes wind instruments)," Technical Report, Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), Paris, 1997.
[5] G. P. Scavone, "Modeling wind instrument sound radiation using digital waveguides," Proceedings of the 1999 International Computer Music Conference, Beijing, China (1999).
[6] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, New York, 1999).
[7] Y. T. Cho, J. S. Bolton, and J. Hald, "Source visualization by using statistically optimized near-field acoustical holography in cylindrical coordinates," J. Acoust. Soc. Am. **118**, 2355–2364 (2005).
[8] S. F. Wu, "On reconstruction of acoustic pressure fields using the Helmholtz equation least squares method," J. Acoust. Soc. Am. **107**, 2511–2522 (2000).
[9] J. Meyers and G. W. Elko, "A highly scalable spherical microphone array based on an orthogonal decomposition of the soundfield," Proceedings of the ICASSP (2002), Vol. **II**, pp. 1781–1784.
[10] B. Rafaely, "Phase-mode versus delay-and-sum spherical microphone array processing," IEEE Signal Process. Lett. **12**, 713–716 (2005).
[11] H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition* (Springer-Verlag, Berlin, 2007).
[12] H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain," IEEE Trans. Acoust., Speech, Signal Process. **34**, 393–398 (1986).
[13] Y. Ando, "On the sound radiation from semi-infinite circular pipe of certain wall thickness," Acustica **22**, 219–225 (1969/1970).
[14] C. Bouchard and D. I. Havelock, "Sound field phase deviations from a point source model for real acoustic sources," Proceedings of the Inter-Noise (2006).

Basic theory and properties of statistically optimized near-field acoustical holography

Jørgen Hald[a)]
Brüel & Kjær Sound & Vibration Measurement A/S, Skodsborgvej 307, DK-2850 Nærum, Denmark

(Received 14 August 2008; revised 19 December 2008; accepted 12 January 2009)

To avoid the requirement set by standard near-field acoustical holography (NAH) to measure an area that fully covers the source, a set of so-called patch NAH methods has been introduced. One such method is the statistically optimized NAH (SONAH). In this method, the acoustic quantities on a mapping surface near the measurement surface are calculated by using a transfer matrix defined in such a way that all propagating waves and a weighted set of evanescent waves are projected with optimal average accuracy. The present paper gives an overview of the basic theory of SONAH, including a description of phenomena such as spatial aliasing and wave-number domain leakage. A revised and generalized mathematical formulation is given, covering the calculation of all three components of particle velocity and the use of up to six virtual source planes. A set of formulas for the inherent estimation error level of the method is derived and used to visualize the regions of validity of the SONAH predictions for some typical microphone array geometries. The sensitivity of the inherent error level distribution to changes in the parameters of the SONAH algorithm is also investigated. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3079773]

PACS number(s): 43.60.Sx, 43.60.Pt, 43.20.Ye, 43.40.Sk [EGW] Pages: 2105–2120

I. INTRODUCTION

Near-field acoustical holography (NAH) is a very useful tool for three-dimensional (3D) visualization of sound radiation and for precise noise source localization based on measurements over a surface near the sound source. Its ability to reconstruct also the evanescent wave components ensures a very high spatial resolution. The oldest form of NAH (Refs. 1–3) is based on regular-grid measurements across a level surface in a separable coordinate system, allowing the NAH calculation to be performed by spatial discrete Fourier transform (DFT). As a consequence, the processing is very fast, but a side effect of using the DFT is severe spatial windowing effects unless the measurement area fully covers the areas with high sound pressure. In some cases this requirement on the measurement area cannot be fulfilled, and in many cases the necessary size becomes prohibitively large.

A set of techniques has therefore been developed to reduce the spatial windowing effects, still maintaining the DFT spatial processing.[4–9] The cost is increased complexity and computational demands. Typically an iterative procedure is first used to extrapolate the measured sound pressure outside the measured area, followed by application of the standard DFT based holography method on the extended data window.

Another set of methods avoiding the use of spatial DFT has also been developed. The decreased computational efficiency can then more or less be compensated by a reduction in the required measurement area. One such method is Helmholtz' equation least-squares (HELS),[10,11] and another similar method is the statistically optimized NAH (SONAH).[12–19] Both use a local model of the sound field in terms of elementary wave functions—typically spherical wave functions for the case of HELS and typically plane wave functions for the case of SONAH—and both perform a (least-squares or least norm) fit of the elementary wave model (EWM) to the measured sound field data. HELS performs the fit of model to data explicitly, while SONAH does it implicitly, as will be detailed in Sec. II A of the present paper. A third much related technique is the equivalent source method.[20–22] Here, the local sound field is represented by a distribution of monopole point sources outside the model region, typically inside the volume of the real source, and a least-squares fit of the monopole excitations to the measured data is performed. The fourth and final technique that shall be mentioned is patch inverse Boundary Element Method (BEM).[22–24]

The SONAH method has been described mostly in conference proceedings[13,15–17] plus three journal papers[12,14,18] that introduce the basic method and a couple of special implementations. A comprehensive description of the basic theory and properties of the method is missing. The aim of the present paper is to give such a description. Some of the ideas from the three conference papers[13,15,17] are used here, but in a common framework that includes a more compact and general mathematical formulation plus a treatment of wave-number leakage. The more general formulation covers, for example, calculation of all three components of particle velocity. Beyond that, the previously described formulations with a single virtual source plane[13] (free-field conditions) and two virtual source planes[15] are extended to the case of six virtual source planes, which allows general sound fields in a box shaped homogeneous regions to be handled. Another new contribution is a description of the inherent error level of the SONAH method and its use to visualize the regions of validity of the sound field predictions.

Results from actual measurements have not been included since that would make the paper too long, but such

a)Electronic mail: jhald@bksv.com

results can be found in Refs. 13–16, 18, and 19. Likewise, no direct comparison of reconstruction accuracies between SONAH and other NAH techniques is given. Such comparisons based on simulated and actual measurements can be found in Refs. 12, 13, 16, and 17.

In Sec. II the basic general theory of the SONAH method is outlined. Two alternative derivations are given in Sec. II A to support the explanation of some important properties of the method. Section II B outlines the planar free-field case of SONAH, and Sec. II C describes the non-free-field cases with two and six virtual source planes. The inherent relative error descriptor of SONAH is then derived in Sec. II D, and Sec. II E introduces the concept of wave-number leakage in relation to SONAH. Section III presents a set of computer simulations to show the accuracy of the method in some typical cases. The focus is on visualizing the regions of validity of the method and on explaining the limitations and the observed errors.

II. THEORY OF SONAH

A. General underlying formulation

Assume that the complex time-harmonic sound pressure $p(\mathbf{r}_i)$ has been measured at a set of positions $\mathbf{r}_i$, $i=1,2,\ldots,I$, in a source-free region Ω occupied by a homogeneous fluid. We consider the problem of reconstructing the sound field in Ω based on the measured pressures. To do that we use a set of elementary wave functions, Ψ_n, $n=1,\ldots,N$, fulfilling the homogeneous wave equation in Ω and capable of representing with a sufficiently high accuracy all sound pressure fields that can exist in Ω. Typically, these elementary wave functions can be plane, cylindrical, or spherical waves, or they can be sound fields from elementary sources such as monopole, dipole, and/or multipole point sources distributed outside Ω. The first step in a typical reconstruction process, as published, for example, in the HELS method,[10,11] is to determine the amplitude and phase of each elementary wave function in the representation of the sound field. This is done by requiring that the EWM represent as accurately as possible the measured pressures:

$$p(\mathbf{r}_i)=\sum_{n=1}^{N} a_n\Psi_n(\mathbf{r}_i), \quad i=1,\ldots,I, \tag{1}$$

a_n being the complex expansion coefficients. If $I\geq N$ and if the measurement points have been suitably chosen, then the expansion coefficients can be found by least-squares solution of Eq. (1). Otherwise the problem is under-determined, and among the infinity of solutions to Eq. (1) the one with the smallest 2-norm will be adopted. The least norm solution will be subsequently referred to as a least norm EWM (LN-EWM), while the least-squares solution shall be called a least-squares EWM (LS-EWM).

To obtain a matrix formulation, we define the matrix $\mathbf{B}$ of wave function values at the measurement positions:

$$\mathbf{B}\equiv\{\Psi_n(\mathbf{r}_i)\}\equiv\begin{Bmatrix}\Psi_1(\mathbf{r}_1) & \Psi_2(\mathbf{r}_1) & \cdots & \Psi_N(\mathbf{r}_1)\\ \Psi_1(\mathbf{r}_2) & \Psi_2(\mathbf{r}_2) & & \Psi_N(\mathbf{r}_2)\\ \vdots & & \ddots & \vdots\\ \Psi_1(\mathbf{r}_I) & \Psi_2(\mathbf{r}_I) & \cdots & \Psi_N(\mathbf{r}_I)\end{Bmatrix} \tag{2}$$

and the vectors $\mathbf{p}$ and $\mathbf{a}$ with measured pressures and with expansion coefficients, respectively:

$$\mathbf{p}\equiv\{p(\mathbf{r}_i)\}\equiv\begin{Bmatrix}p(\mathbf{r}_1)\\ p(\mathbf{r}_2)\\ \vdots\\ p(\mathbf{r}_I)\end{Bmatrix}, \quad \mathbf{a}\equiv\{a_n\}\equiv\begin{Bmatrix}a_1\\ a_2\\ \vdots\\ a_N\end{Bmatrix}. \tag{3}$$

Using these definitions Eq. (1) can be rewritten in the following form:

$$\mathbf{Ba}=\mathbf{p}. \tag{4}$$

With $I\geq N$ the least-squares (LS-EWM) solution is $\mathbf{a}=(\mathbf{B}^H\mathbf{B})^{-1}\mathbf{B}^H\mathbf{p}$, and with $I<N$ the least norm (LN-EWM) solution is[25] $\mathbf{a}=\mathbf{B}^H(\mathbf{BB}^H)^{-1}\mathbf{p}$, the symbol H representing Hermitian transpose. In practice, regularization has to be applied in these solutions. We shall consider only Tikhonov regularization, and in that case the LS-EWM and LN-EWM formulas become

$$\tilde{\mathbf{a}}=(\mathbf{B}^H\mathbf{B}+\varepsilon\mathbf{I})^{-1}\mathbf{B}^H\mathbf{p} \tag{5}$$

and

$$\tilde{\mathbf{a}}=\mathbf{B}^H(\mathbf{BB}^H+\varepsilon\mathbf{I})^{-1}\mathbf{p}, \tag{6}$$

respectively, ε being a positive regularization parameter and $\mathbf{I}$ a unit diagonal quadratic matrix of appropriate dimensions. It is shown in Appendix A that actually the two regularized formulas (5) and (6) provide exactly the same solution independent of the values of I and N. But for $I<N$ it is still a computational advantage to use the LN-EWM formula (6) because the matrix $\mathbf{BB}^H$ has smaller dimension than $\mathbf{B}^H\mathbf{B}$.

The combined least-squares and least norm properties of, for example, the regularized least-squares formula (5) are evident from the following statement given by Hansen:[26] The solution $\tilde{\mathbf{a}}$ provided by Eq. (5) with any positive regularization parameter ε has the smallest norm $\|\mathbf{a}\|_2$ among all those coefficient vectors $\mathbf{a}$ that provide residual norms not exceeding $\|\mathbf{B}\tilde{\mathbf{a}}-\mathbf{p}\|_2$, and it has the smallest residual norm $\|\mathbf{Ba}-\mathbf{p}\|_2$ among all vectors $\mathbf{a}$ with a norm that does not exceed $\|\tilde{\mathbf{a}}\|_2$.

Assuming from now on that the elementary wave functions Ψ_n constitute a complete basis in Ω, all sound pressure fields can be represented by these functions, so we can use the vector $\mathbf{a}$ of expansion coefficients to represent the true pressure in Ω. Plane wave spectra and spherical wave functions constitute complete basis function sets in suitably defined domains. The vector $\tilde{\mathbf{a}}$ then represents the corresponding coefficient vector estimated from the array measurements. Using that vector $\tilde{\mathbf{a}}$, the reconstructed sound pressure $\tilde{p}(\mathbf{r})$ at an arbitrary point $\mathbf{r}$ in Ω can be calculated by summation of the wave function contributions:

$$\tilde{p}(\mathbf{r}) \equiv \sum_{n=1}^{N} \tilde{a}_n \Psi_n(\mathbf{r}) = \tilde{\mathbf{a}}^T \boldsymbol{\alpha}(\mathbf{r}), \quad \boldsymbol{\alpha}(\mathbf{r}) \equiv \begin{Bmatrix} \Psi_1(\mathbf{r}) \\ \Psi_2(\mathbf{r}) \\ \vdots \\ \Psi_N(\mathbf{r}) \end{Bmatrix}, \tag{7}$$

T representing the transpose of a matrix or vector. This reconstructed pressure field fulfils the homogeneous wave equation in Ω because all expansion functions do. The particle velocity in a direction χ can then be estimated from the reconstructed sound pressure by application of Euler's equation:

$$\tilde{u}_\chi(\mathbf{r}) = \frac{-1}{j\omega\rho_0} \frac{\partial \tilde{p}(\mathbf{r})}{\partial \chi}. \tag{8}$$

Here, ω is the angular frequency, ρ_0 is the density of the medium, and the implicit time variation is $e^{j\omega t}$. Other symbols to be used subsequently are the propagation speed of sound, c, and the wave-number, $k \equiv \omega/c$.

The LS-EWM method described above applied to a single-origin spherical wave representation of the sound field is a basic version of the HELS method. The region of validity, Ω, of the field model is in that case a finite or infinite homogeneous spherical shell centered at the origin of the spherical wave functions. Spherical wave functions are efficient for representing the sound field in such regions because they are discretely indexed and because a finite subset can provide an accurate representation in a finite spherical shell. Cylindrical and plane waves by nature constitute continuous spectra in a wave-number domain, so use of these in the above formulation must involve a subset of the wave functions obtained by sampling in the wave-number domain. The SONAH method to be introduced now overcomes this sampling problem by solving the field estimation problem in a dual way.

Following the earlier derivations by Hald[13,15] of SONAH based on plane wave functions, we require the reconstructed sound pressure $\tilde{p}(\mathbf{r})$ at a point $\mathbf{r}$ in Ω to be obtained as a linear combination of the known sound pressure values:

$$\tilde{p}(\mathbf{r}) = \mathbf{p}^T \mathbf{c}(\mathbf{r}), \tag{9}$$

where the vector $\mathbf{c}(\mathbf{r})$ contains the complex estimation weights. The vector $\mathbf{c}$ is then obtained by requiring that the prediction formula (9) provides in a least-squares sense the best average prediction for all expansion functions Ψ_n. With reference to Eqs. (2) and (7), this means that $\mathbf{c}(\mathbf{r})$ should be found as the least-squares solution to the following set of linear equations:

$$\mathbf{B}^T \mathbf{c}(\mathbf{r}) = \boldsymbol{\alpha}(\mathbf{r}). \tag{10}$$

Clearly, this equation is under-determined whenever the EWM Eq. (4) is over-determined and vice versa. Only the Tikhonov regularized solution shall be considered, meaning that the least-squares and least norm solutions to Eq. (10) can be both expressed in terms of the least-squares formula:

$$\mathbf{c}(\mathbf{r}) = (\mathbf{A}^H \mathbf{A} + \varepsilon \mathbf{I})^{-1} \mathbf{A}^H \boldsymbol{\alpha}(\mathbf{r}), \tag{11}$$

where the matrix $\mathbf{A}$ is the transpose of $\mathbf{B}$:

$$\mathbf{A} \equiv \mathbf{B}^T. \tag{12}$$

From the definitions (7) and (12), it follows that the elements of $\mathbf{A}^H\mathbf{A}$ and $\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})$ are given by formulas (13) and (14) below, respectively, * representing complex conjugate:

$$[\mathbf{A}^H \mathbf{A}]_{ij} = \sum_n \Psi_n^*(\mathbf{r}_i) \Psi_n(\mathbf{r}_j), \tag{13}$$

$$[\mathbf{A}^H \boldsymbol{\alpha}(\mathbf{r})]_i = \sum_n \Psi_n^*(\mathbf{r}_i) \Psi_n(\mathbf{r}). \tag{14}$$

The matrix $\mathbf{A}^H\mathbf{A}$ can be seen as a matrix of cross correlations between the measurement points in the domain of the elementary wave functions (expansion functions), and the vector $\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})$ then contains the cross correlations between the measurement points and the estimation position $\mathbf{r}$ (see Steiner and Hald[12]). Notice that $\mathbf{A}^H\mathbf{A}$ is a Hermitian symmetric matrix with non-negative eigenvalues, and that its dimensions equal the number I of measurement points, independent of the number of expansion functions. This will be very effectively exploited in connection with continuous spectra of wave functions.

The SONAH prediction formulation just derived can be shown to provide the same solution as the LS-EWM and LN-EWM formulas. Actually, the LN-EWM Eqs. (7) and (6) easily lead to the SONAH formulation:

$$\begin{aligned} \tilde{p}(\mathbf{r}) = \tilde{\mathbf{a}}^T \boldsymbol{\alpha}(\mathbf{r}) &= \mathbf{p}^T [(\mathbf{B}\mathbf{B}^H + \varepsilon \mathbf{I})^{-1}]^T \mathbf{B}^* \boldsymbol{\alpha}(\mathbf{r}) \\ &= \mathbf{p}^T [(\mathbf{B}\mathbf{B}^H + \varepsilon \mathbf{I})^T]^{-1} \mathbf{B}^* \boldsymbol{\alpha}(\mathbf{r}) \\ &= \mathbf{p}^T [\mathbf{A}^H \mathbf{A} + \varepsilon \mathbf{I}]^{-1} \mathbf{A}^H \boldsymbol{\alpha}(\mathbf{r}) = \mathbf{p}^T \mathbf{c}(\mathbf{r}). \end{aligned} \tag{15}$$

Here we have used the fact that for any non-singular quadratic matrix the transpose of the inverse equals the inverse of the transpose, and we have used Eqs. (12) and (11).

In EWM methods, where a vector $\tilde{\mathbf{a}}$ of expansion coefficients is explicitly calculated, spatial filtering/smoothing can be done both in the regularized solution for $\tilde{\mathbf{a}}$ and by attenuating the components of $\tilde{\mathbf{a}}$ representing high spatial frequencies before $\tilde{\mathbf{a}}$ is used for sound field reconstructions. Since SONAH does not calculate $\tilde{\mathbf{a}}$, spatial filtering/smoothing can be done only in the regularization. Making this role of the regularization work properly, however, requires a suitable scaling of the elementary wave functions, as will be explained in Sec. II B focusing on plane wave representations.

An interesting difference between the presented EWM and SONAH derivations is that the SONAH derivation considers from the beginning with Eqs. (9) and (10) the prediction problem, whereas the EWM derivation just fits a set of basis functions at the measurement positions.

B. SONAH plane wave formulation for free-field conditions

We consider now the special case where the elementary waves are plane propagating and evanescent waves, and we start with free-field conditions, as illustrated in Fig. 1, where sources behind a source plane, $z=z_0^+$, $z_0^+<0$, radiate into the

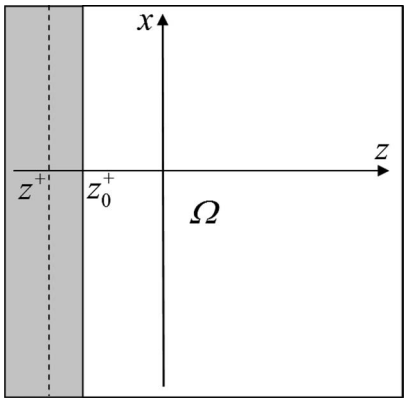


FIG. 1. Free-field measurement geometry with only one source plane (vertical line) and one virtual source plane (dashed vertical line). The domain Ω of the sound field representation is white.

source free, homogeneous half space Ω with $z \geq z_0^+$. In this half space, the radiated sound field can be represented by plane wave functions of the form

$$\Phi_{\mathbf{k}}(\mathbf{r}) \equiv F(k_z) e^{-j(k_x x + k_y y + k_z (z - z^+))}, \tag{16}$$

where F is an amplitude weighting function, $z=z^+$ is the so-called virtual source plane where the wave functions are scaled, $\mathbf{r} \equiv (x,y,z)$ is a position vector, and $\mathbf{k} \equiv (k_x, k_y, k_z)$ is a wave-number vector with

$$k_z \equiv \begin{cases} \sqrt{k^2 - (k_x^2 + k_y^2)} & (k_x^2 + k_y^2) \leq k^2 \\ -j\sqrt{(k_x^2 + k_y^2) - k^2} & (k_x^2 + k_y^2) > k^2 \end{cases}. \tag{17}$$

Thus, the complete set of expansion functions constitutes a continuous spectrum in the two wave-number variables k_x and k_y. The virtual source plane $z=z^+$ could be coincident with the real source plane, $z=z_0^+$, but typically it should be chosen behind that plane, i.e., $z^+ < z_0^+$. The superscript + on the coordinates of the real and virtual source planes indicates that waves propagating in the positive z-axis direction are represented.

To use the EWM or SONAH formulations given in the previous section, we select first a discrete subset by regular sampling with intervals $(\Delta k_x, \Delta k_y)$ in the two-dimensional (2D) wave-number domain (k_x, k_y):

$$\Psi_n(\mathbf{r}) \equiv \kappa \Phi_{\mathbf{k}_n}(\mathbf{r}), \quad \kappa \equiv \sqrt{\frac{\Delta k_x \Delta k_y}{2\pi k^2}}. \tag{18}$$

Here, $\mathbf{k}_n$ are the wave-number domain sampling points, and the scaling factor κ has been introduced to support a smooth transition to the use of continuous wave-number spectra. When choosing the sampling spacing, one should be aware that a regular discrete representation in the wave-number domain will represent a field that is spatially periodic,[3] leading to wrap-around errors just like in traditional NAH based on spatial DFT processing. The wave-number domain sampling spacing should therefore be smaller than 2π divided by the smallest acceptable distance between the periodic replica in the represented sound field. At the same time, the sampling area should cover all propagating waves and all those evanescent waves that have significant amplitude in the measurement region. This almost inevitably leads to a number N of expansion functions Ψ_n larger than the number I of measurement positions. The equivalent EWM problem will therefore be under-determined, and the solution (6) will be dominantly a least norm solution.

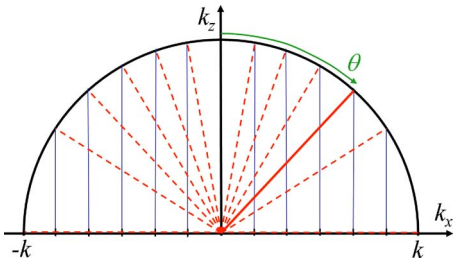


FIG. 2. (Color online) Equidistant samples along the k_x-axis and their mapping on corresponding directions of plane wave propagation.

We consider first a unit weighting, $F(k_z)=1$, of the wave functions, as used in previous publications on SONAH. Then according to Eq. (16) all plane wave functions $\Phi_{\mathbf{k}}(\mathbf{r})$ have amplitude equal to 1 on the virtual source plane, $z=z^+$, and all elements in the sampled and scaled discrete subset $\Psi_n(\mathbf{r})$ have amplitude equal to κ on that plane. The EWM expansion coefficient vector, $\tilde{\mathbf{a}}$, is therefore a scaled discrete wave-number spectrum of the reconstructed sound pressure on the virtual source plane, and the matrix $\mathbf{B}$ can be seen as a matrix of transfer functions from that scaled plane wave spectrum to the sound pressure values at the microphone positions. According to the definition in Eq. (2), the elements of $\mathbf{B}$ are given by Eq. (18) with the microphone coordinates inserted. Clearly, through the transfer matrix $\mathbf{B}$, evanescent wave components ($\Phi_{\mathbf{k}}$ with $k_x^2 + k_y^2 > k^2$) will get an attenuation that increases exponentially with wave-number. The overall trend must therefore be that the higher wave-number components will be represented in the smaller singular values of the matrix $\mathbf{B}$. This property ensures that appropriate filtering of the evanescent wave components can be performed by Tikhonov regularization in the inverse solution for the scaled wave-number spectrum, $\tilde{\mathbf{a}}$. Some combinations of propagating waves will also be represented by small singular values, for example, combinations that represent sound radiation from areas in the virtual source plane far from the measurement positions.

Because of the above mentioned under-determined nature of the EWM problem, the solution will dominantly be a minimum norm solution. A further reduction in the norm of the solution vector, $\tilde{\mathbf{a}}$, will be introduced by the regularization. The solution process will therefore minimize the energy in the wave-number spectrum and consequently also, according to Parseval's theorem,[27] the energy in the pressure distribution on the virtual source plane (and on parallel planes) under the constraint that the pressure at the microphones is well represented. As a result, a good sound field representation is achieved within a volume around the array, while an increasing underestimation will be seen away from the array. Since SONAH and EWM provide the same reconstruction [see Eq. (15)], the consequences will be the same for SONAH. The simulated measurements to be presented in Sec. III confirm this hypothesis.

Figure 2 illustrates the regular sampling of Eq. (18)

along the k_x wave-number axis at $k_y=0$. Inside the radiation circle, i.e., for $k_x^2+k_y^2 \le k^2$, a specific sample at (k_x,k_y) corresponds to a plane wave propagating in the direction of the vector (k_x,k_y,k_z), with k_z given by Eq. (17).[3] In Fig. 2, this direction can be found by projecting the sample position k_x up on the shown half circle with radius k. Clearly, with a constant density of the sample points in the (k_x,k_y)-plane, the angular density of the plane waves will be highest close to the z-direction and will decrease when the off-axis angle θ increases toward 90°. Thus, if we use a constant weighting function F, then plane waves propagating near the axial direction will get a higher weight than the far off-axis waves in the least-squares solution (11) for the prediction weights of SONAH. This can be compensated by applying a higher weight F to the off-axis waves. Considering an infinitely small area segment on the (k_x,k_y)-plane, it will represent an area on the directional hemisphere $k_x^2+k_y^2+k_z^2=k^2$, which is larger by a factor of $1/\cos(\theta)=k/|k_z|$. A constant directional power density of the weighting function can therefore be obtained by choosing F such that $|F(k_z)|^2 k/|k_z|$ is constant, for example,

$$F(k_z) = F_0 \sqrt{\frac{k}{|k_z|}}, \tag{19}$$

where F_0 is a constant. Equations (13), (14), (16), and (18) show that the contributions to the elements of the matrices $\mathbf{A}^H\mathbf{A}$ and $\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})$ are obtained by summation over wave-number contributions, which are scaled by $|F|^2$. This is the reason for choosing the equalization to be on the directional power density and not on the directional amplitude density. The "omni-directional weighting" of Eq. (19), applied also outside the radiation circle, has been found to provide on the average slightly better accuracy than the constant weighting $F(k_z)=1$, and it will be used subsequently as the default weighting.

To remove completely the wrap-around problem caused by the discrete representation in the wave-number domain, we now let the sampling spacing in the (k_x,k_y)-plane go to zero, meaning that the summations in Eqs. (13) and (14) become integrals over the (k_x,k_y)-plane. From Eqs. (18), (13), and (14) we obtain

$$[\mathbf{A}^H\mathbf{A}]_{ij} = \sum_n \Phi_{\mathbf{k}_n}^*(\mathbf{r}_i)\Phi_{\mathbf{k}_n}(\mathbf{r}_j)\frac{\Delta k_x \Delta k_y}{2\pi k^2} \underset{\substack{\Delta k_x\to 0\\ \Delta k_y\to 0}}{\to} \frac{1}{2\pi k^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi_{\mathbf{k}}^*(\mathbf{r}_i)\Phi_{\mathbf{k}}(\mathbf{r}_j)dk_x dk_y, \tag{20}$$

$$[\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})]_i = \sum_n \Phi_{\mathbf{k}_n}^*(\mathbf{r}_i)\Phi_{\mathbf{k}_n}(\mathbf{r})\frac{\Delta k_x \Delta k_y}{2\pi k^2} \underset{\substack{\Delta k_x\to 0\\ \Delta k_y\to 0}}{\to} \frac{1}{2\pi k^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi_{\mathbf{k}}^*(\mathbf{r}_i)\Phi_{\mathbf{k}}(\mathbf{r})dk_x dk_y. \tag{21}$$

The $1/2\pi k^2$ scaling factor has been chosen with the aim of obtaining nice dimensionless expression for the matrix elements. Letting the sampling spacing go to zero will tend to strengthen the under-determined nature of the problem, motivating further the numerical investigations of the estimation errors in Sec. III. These investigations show that very good accuracy is, in general, achieved within a volume around the measurement positions, and this conclusion agrees with results from both simulated and actual measurements.[12–19] In particular, numerous published[12,13] and un-published applications to small measurement grids (patch calculations) show a significant gain in reconstruction accuracy by replacing the wrap-around errors of DFT based NAH with the least norm errors of plane wave SONAH.

An interesting observation in relation to Eqs. (20) and (21) above is that the discrete wave model can be seen as a way of performing numerical computation of the integrals, with the same function evaluation points (Gauss points) used in all integrals.

Use of the limiting values (20) and (21) of the matrix elements implies that the matrix $\mathbf{A}$ and the vector $\boldsymbol{\alpha}(\mathbf{r})$ have a dimension that is infinite. $\mathbf{A}$ and $\boldsymbol{\alpha}(\mathbf{r})$ will be used anyway in a couple of subsequent derivations, where they should be thought of as limiting cases for $N\to\infty$, corresponding to the sampling intervals $(\Delta k_x,\Delta k_y)$ approaching zero. The vector $\tilde{\mathbf{a}}$ of expansion coefficients of the equivalent EWM should in the same way be considered as a limiting case.

As shown in Appendix B, the 2D integrals in Eqs. (20) and (21) can be reduced to one-dimensional integrals through application of Eqs. (16), (17), and (19): The integrals can be expressed in terms of the following two functions of two variables:

$$I_p(\rho,\zeta) \equiv \int_0^{\pi/2} J_0(\rho\sin(\theta))e^{j\zeta\cos(\theta)}\sin(\theta)d\theta, \tag{22}$$

$$II_p(\rho,\zeta) \equiv \int_0^{\infty} J_0(\rho\sqrt{1+v^2})e^{-\zeta v}dv, \tag{23}$$

where J_0 is the Bessel function of order zero. The function I_p represents contributions from inside the radiation circle in the wave-number domain, and the function II_p represents contributions from outside the radiation circle. The contributions I_p and II_p can be combined into a single function III_p defined as

$$III_p(\rho,\zeta_1,\zeta_2) \equiv I_p(\rho,\zeta_1) + II_p(\rho,\zeta_2). \tag{24}$$

Using this function, formulas (20) and (21) for the matrix elements can be reformulated as follows (see Appendix B):

$$[\mathbf{A}^H\mathbf{A}]_{ij} = F_0^2 III_p(kR_{ij}^{xy}, k(z_i-z_j), k(z_i+z_j-2z^+)), \tag{25}$$

$$[\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})]_i = F_0^2 III_p(kR_i^{xy}, k(z_i-z), k(z_i+z-2z^+)). \tag{26}$$

Here, $\mathbf{r}_i \equiv (x_i,y_i,z_i)$ and $\mathbf{r}_j \equiv (x_j,y_j,z_j)$ are two microphone positions, the pressure estimation position is $\mathbf{r}\equiv(x,y,z)$, R_{ij}^{xy} is the distance between the projections of microphone positions i and j on the xy-plane (or, more generally, on the virtual source plane), and R_i^{xy} is the distance between the projections of microphone position i and the pressure calculation position on the xy-plane.

Different methods for determination of the regularization parameter ε of Eq. (11) have been investigated by Go-

mes and Hansen,[16] and the *L*-curve method was found to provide the overall most stable result in connection with SONAH. Earlier studies with uncorrelated measurement errors had pointed at generalized cross validation (GCV). The second important parameter in SONAH is the position $z=z^+$ of the virtual source plane. Results published by Hald and Gomes[17] show that optimal accuracy is obtained by positioning the virtual source plane one to one and a half measurement grid spacing behind the physical source plane, i.e., further away from the measurement grid than the physical source surface. Slightly larger distances were often better at low frequencies. In case of an irregular measurement grid, the criterion should use the average grid spacing.

To obtain particle velocity estimates, we apply Euler's Eq. (8) to the SONAH pressure estimate given by Eqs. (9) and (11). As a result we obtain

$$\tilde{u}_\chi(\mathbf{r}) = \frac{-1}{j\omega\rho_0}\frac{\partial\tilde{p}(\mathbf{r})}{\partial\chi} = \frac{1}{\rho_0 c}\mathbf{p}^T(\mathbf{A}^H\mathbf{A}+\varepsilon\mathbf{I})^{-1}\mathbf{A}^H\boldsymbol{\beta}_\chi(\mathbf{r}), \quad (27)$$

where

$$\mathbf{A}^H\boldsymbol{\beta}_\chi(\mathbf{r}) \equiv \frac{-1}{jk}\frac{\partial\{\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})\}}{\partial\chi}. \quad (28)$$

Explicit formulas for $\mathbf{A}^H\boldsymbol{\beta}_\chi(\mathbf{r})$ can be obtained by inserting Eq. (26) in Eq. (28), followed by application of expressions (22) and (23) for the functions I_p and II_p. For the z-component we obtain

$$[\mathbf{A}^H\boldsymbol{\beta}_z(\mathbf{r})]_i = F_0^2 III_z(kR_i^{xy}, k(z_i-z), k(z_i+z-2z^+)), \quad (29)$$

and for the x- and y-components we obtain

$$\frac{[\mathbf{A}^H\boldsymbol{\beta}_x(\mathbf{r})]_i}{jk(x_i-x)} = \frac{[\mathbf{A}^H\boldsymbol{\beta}_y(\mathbf{r})]_i}{jk(y_i-y)} = F_0^2 III_{xy}(kR_i^{xy}, k(z_i-z), k(z_i+z-2z^+)). \quad (30)$$

The new functions III_z and III_{xy} are listed in Appendix C together with III_p. Appendix C also gives the form these functions would have with a constant weighting ($F(k_z)=1$) of the elementary waves, and some considerations are given to the calculation of the functions.

Typically in the present free-field case a planar array will be used, but SONAH supports in any case arbitrary 3D measurement grids.

C. SONAH plane wave formulation for non-free-field conditions

When sources and reflecting objects are not restricted to a half space, then the plane wave functions defined in Eq. (16) no longer constitute a complete basis for sound fields in the measurement region.

We consider first the case illustrated in Fig. 3, where measurements are taken in a homogeneous and source-free slab/slice between two parallel source planes, $z_0^+ \le z \le z_0^-$. The part of the sound field with sources for $z<z_0^+$ can be represented by the basis functions of Eq. (16), to be written in the following as $\Phi_{\mathbf{k}}^{z+}$. To represent also the sound field components with sources for $z>z_0^-$, we have to add a corresponding set of plane wave functions $\Phi_{\mathbf{k}}^{z-}$ with virtual source plane at $z=z^-$, $z^- \ge z_0^-$, resulting in the combined set

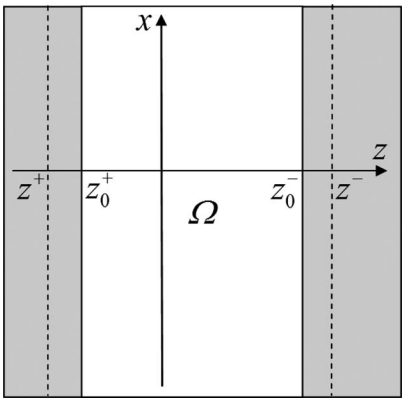


FIG. 3. Two parallel source planes limiting a homogeneous and source-free region. The two corresponding virtual source planes are shown as dashed lines.

$$\Phi_{\mathbf{k}}^{z+}(\mathbf{r}) \equiv F_{z+}(k_z)e^{-j(k_x x+k_y y+k_z(z-z^+))}, \quad (31)$$

$$\Phi_{\mathbf{k}}^{z-}(\mathbf{r}) \equiv F_{z-}(k_z)e^{-j(k_x x+k_y y-k_z(z-z^-))}, \quad (32)$$

where the definition of k_z in Eq. (17) is still used, and omnidirectional weighting is applied for both sets:

$$F_{z+}(k_z) = F_{0,z+}\sqrt{\frac{k}{|k_z|}}, \quad F_{z-}(k_z) = F_{0,z-}\sqrt{\frac{k}{|k_z|}}. \quad (33)$$

A set of plane wave functions as given in Eq. (31) or Eq. (32) is often referred to just as a "virtual source plane" in the present paper. So adding a virtual source plane means adding a set of plane wave functions in the SONAH (or EWM) method. Equations (13), (14), (20), and (21) show that a SONAH estimation procedure taking into account the combined set of basis functions, Eqs. (31) and (32), can be obtained by adding the contributions from the two sets in the elements of the matrices $\mathbf{A}^H\mathbf{A}$ and $\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})$. The result of doing this is

$$[\mathbf{A}^H\mathbf{A}]_{ij} = F_{0,z+}^2 III_p(kR_{ij}^{xy}, k(z_i-z_j), k(z_i+z_j-2z^+)) + F_{0,z-}^2 III_p(kR_{ij}^{xy}, k(z_j-z_i), k(2z^- - z_i - z_j)), \quad (34)$$

$$[\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})]_i = F_{0,z+}^2 III_p(kR_i^{xy}, k(z_i-z), k(z_i+z-2z^+)) + F_{0,z-}^2 III_p(kR_i^{xy}, k(z-z_i), k(2z^- - z_i - z)). \quad (35)$$

The vectors $\mathbf{A}^H\boldsymbol{\beta}_\chi(\mathbf{r})$ needed for estimation of particle velocity can be derived just like for the free-field case with the following result:

$$[\mathbf{A}^H\boldsymbol{\beta}_z(\mathbf{r})]_i = F_{0,z+}^2 III_z(kR_i^{xy}, k(z_i-z), k(z_i+z-2z^+)) - F_{0,z-}^2 III_z(kR_i^{xy}, k(z-z_i), k(2z^- - z_i - z)), \quad (36)$$

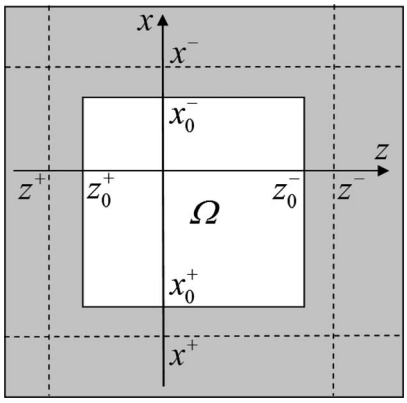


FIG. 4. Source planes on all sides of a box shaped homogeneous and source-free region. The corresponding virtual source planes are shown as dashed lines.

$$\frac{[\mathbf{A}^H\boldsymbol{\beta}_x(\mathbf{r})]_i}{jk(x_i-x)}=\frac{[\mathbf{A}^H\boldsymbol{\beta}_y(\mathbf{r})]_i}{jk(y_i-y)}$$
$$=F_{0,z+}^2 III_{xy}(kR_i^{xy},k(z_i-z),k(z_i+z-2z^+))$$
$$+F_{0,z-}^2 III_{xy}(kR_i^{xy},k(z-z_i),k(2z^--z_i-z)). \quad (37)$$

The above formulation puts no restriction on the selection of the measurement positions, but a planar array will not be sufficient to get accurate predictions. This can be illustrated by the fact that sources which are image symmetric in the measurement plane cannot then be distinguished because at every microphone they produce the same pressure. Typically, a dual layer array consisting of two identical planar rectangular arrays will be used as described by Hald.[15] The weight on the waves with sources on either side of the array can be adjusted through the coefficients $F_{0,z+}$ and $F_{0,z-}$. Typically, both parameters will be chosen equal to 1, but if no significant sources are present on one side of the array, better estimates can be achieved by setting one of them equal to zero, providing thereby the information of the single-sided source distribution to the formulation. In the applications of the present paper, the coefficient are set to 1.

The extension of the case with sources on both sides of an infinite planar slab/slice to the case illustrated in Fig. 4 with sources on all six sides of a box shaped volume is straightforward: We just add contributions to the matrix elements from six sets of plane wave functions belonging to the respective six sides of the box. Actually, with omnidirectional weighting of the plane wave functions, the formulas for sources on only two opposite sides take into account plane propagating waves from all directions because each side includes all plane waves from a hemisphere with equal weights. But the two sides do not include all types of evanescent waves: Waves decaying in the x-and y-directions are missing.

D. Generic inherent estimation errors

According to the derivation in Sec. II A, SONAH performs a linear prediction of a sound field parameter at a calculation point based on a set of measured quantities, the prediction coefficients being characterized as providing in a least-squares sense the best estimate for all plane propagating waves and a weighted set of the evanescent waves. It is then interesting to see how well on an average the values of these wave functions are predicted at the calculation point.

Considering first free-field conditions, the set of plane propagating and evanescent wave functions is given by Eq. (18) with the number of functions approaching infinity as described in Eqs. (20) and (21). The vector $\boldsymbol{\alpha}(\mathbf{r})$ contains according to the definition in Eq. (7) the true values of all these elementary waves at the calculation point $\mathbf{r}$. Equations (18), (12), and (2) show that one row of the matrix $\mathbf{A}$ contains the values of one of the wave functions across the measurement positions. In the process of calculating the prediction error, we need a possibility of using a spectrum of test wave functions different from the spectrum used in the calculation of the prediction weight vector $\mathbf{c}(\mathbf{r})$ in Eq. (11). Here, we will consider only the case where the test wave function spectrum has also the form given in Eq. (16) but is scaled in the plane $z=z_s^+$ instead of the virtual source plane $z=z^+$ used in the SONAH prediction. This new virtual source plane, $z=z_s^+$, can be used to set up the matrix $\mathbf{A}_s$ and the vector $\boldsymbol{\alpha}_s(\mathbf{r})$. Since one row of $\mathbf{A}_s$ contains the values of one of the test wave functions across all measurement positions, it follows from Eq. (9) that the vector $\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})$ defined as

$$\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})\equiv\mathbf{A}_s\mathbf{c}(\mathbf{r}) \quad (38)$$

contains the SONAH estimates of the test wave functions at the calculation point. The 2-norm of this vector is a measure of the average amplitude of the predictions:

$$\|\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})\|_2^2=\tilde{\boldsymbol{\alpha}}_s^H\tilde{\boldsymbol{\alpha}}_s=\mathbf{c}^H(\mathbf{r})[\mathbf{A}_s^H\mathbf{A}_s]\mathbf{c}(\mathbf{r}). \quad (39)$$

Here, the matrix $\mathbf{A}_s^H\mathbf{A}_s$ can be calculated in the same way as the matrix $\mathbf{A}^H\mathbf{A}$, just using the virtual source plane $z=z_s^+$. An average underestimation will be characterized by the prediction norm $\|\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})\|_2$ being smaller than the true norm $\|\boldsymbol{\alpha}_s(\mathbf{r})\|_2$. To find the true norm $\|\boldsymbol{\alpha}_s(\mathbf{r})\|_2$, we notice from Eqs. (18), (12), and (2) that one column of the matrix $\mathbf{A}_s$ contains the exact values of all test wave functions at a microphone position, just like $\boldsymbol{\alpha}_s(\mathbf{r})$ contains all function values at the calculation point. The value of $\boldsymbol{\alpha}_s^H\boldsymbol{\alpha}_s$ can therefore be calculated using expression (25) for a diagonal element $[\mathbf{A}_s^H\mathbf{A}_s]_{ii}$ of the matrix $\mathbf{A}_s^H\mathbf{A}_s$, just by replacing the microphone position $\mathbf{r}_i$ by the calculation position $\mathbf{r}$:

$$\|\boldsymbol{\alpha}_s(\mathbf{r})\|_2^2=\boldsymbol{\alpha}_s^H\boldsymbol{\alpha}_s=F_0^2 I_p(0,0)+F_0^2 II_p(0,k(2z-2z_s^+))$$
$$=F_0^2\left[1+\frac{1}{2k(z-z_s^+)}\right]. \quad (40)$$

Here, the analytic expressions for the two integrals can be easily verified using the definitions (22) and (23). The inherent average amplitude gain $A(\mathbf{r})$ is then defined as

$$A(\mathbf{r})\equiv\frac{\|\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})\|_2}{\|\boldsymbol{\alpha}_s(\mathbf{r})\|_2}. \quad (41)$$

The 2-norm of the estimation error vector $\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})-\boldsymbol{\alpha}_s(\mathbf{r})$ can be calculated by application of Eq. (38):

$$\|\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})-\boldsymbol{\alpha}_s(\mathbf{r})\|_2^2=(\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})-\boldsymbol{\alpha}_s(\mathbf{r}))^H(\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})-\boldsymbol{\alpha}_s(\mathbf{r}))$$
$$=\|\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})\|_2^2+\|\boldsymbol{\alpha}_s(\mathbf{r})\|_2^2$$
$$-2\,\mathrm{Re}\{\boldsymbol{\alpha}_s^H(\mathbf{r})\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})\}=\|\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})\|_2^2$$
$$+\|\boldsymbol{\alpha}_s(\mathbf{r})\|_2^2-2\,\mathrm{Re}\{[\mathbf{A}_s^H\boldsymbol{\alpha}_s(\mathbf{r})]^H\mathbf{c}(\mathbf{r})\}, \tag{42}$$

and finally the generic inherent average estimation error level $E(\mathbf{r})$ is defined as

$$E(\mathbf{r})\equiv 10\log\left(\frac{\|\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})-\boldsymbol{\alpha}_s(\mathbf{r})\|_2^2}{\|\boldsymbol{\alpha}_s(\mathbf{r})\|_2^2}\right). \tag{43}$$

When the wave-number spectrum of the test field equals the spectrum used in the SONAH derivation, i.e., $\boldsymbol{\alpha}_s(\mathbf{r})=\boldsymbol{\alpha}(\mathbf{r})$, then because of Eqs. (10) and (38) the numerator in Eq. (43) is just the residual in the solution of Eq. (10) for the SONAH prediction weight vector.

The above description can be extended to non-free-field conditions simply by application of the extended formulas for the elements of the matrices $\mathbf{A}_s^H\mathbf{A}_s$ and $\mathbf{A}_s^H\boldsymbol{\alpha}_s(\mathbf{r})$, for example, Eqs. (34) and (35) for the case of sources on two sides. In that case, expression (40) for $\boldsymbol{\alpha}_s^H\boldsymbol{\alpha}_s$ has to be extended just like Eq. (25) was extended to Eq. (34):

$$\boldsymbol{\alpha}_s^H\boldsymbol{\alpha}_s=F_{0,z+}^2 I_p(0,0)+F_{0,z+}^2 II_p(0,k(2z-2z_s^+))$$
$$+F_{0,z-}^2 I_p(0,0)+F_{0,z-}^2 II_p(0,k(2z_s^--2z))$$
$$=F_{0,z+}^2\left[1+\frac{1}{2k(z-z_s^+)}\right]+F_{0,z-}^2\left[1+\frac{1}{2k(z_s^--z)}\right]. \tag{44}$$

The question to be considered next is how the generic inherent error level relates to the sound pressure prediction error for a particular sound field. To do this we consider a test sound field specified by its true plane wave expansion coefficient vector $\mathbf{a}$, and we allow for different virtual source planes for the test waves and for the SONAH/EWM predictions. In analogy with Eq. (7), the sound pressure at the calculation position is then $p(\mathbf{r})=\mathbf{a}^T\boldsymbol{\alpha}_s(\mathbf{r})$, and similarly from Eqs. (4) and (12) the vector of sound pressure values at the microphones is $\mathbf{p}=\mathbf{A}_s^T\mathbf{a}$. Use of this second relation in combination with Eq. (15) leads to the following expression for the EWM expansion coefficient vector estimate:

$$\tilde{\mathbf{a}}^T=\mathbf{p}^T[\mathbf{A}^H\mathbf{A}+\varepsilon\mathbf{I}]^{-1}\mathbf{A}^H=\mathbf{a}^T\mathbf{A}_s[\mathbf{A}^H\mathbf{A}+\varepsilon\mathbf{I}]^{-1}\mathbf{A}^H, \tag{45}$$

and the estimated sound pressure is therefore $\tilde{p}(\mathbf{r})=\tilde{\mathbf{a}}^T\boldsymbol{\alpha}(\mathbf{r})$. To rewrite that in terms of the estimates $\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})$ of the test plane wave functions (based on their values at the microphones), we use the following expression obtained from Eqs. (38) and (11):

$$\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})=\mathbf{A}_s[\mathbf{A}^H\mathbf{A}+\varepsilon\mathbf{I}]^{-1}\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r}). \tag{46}$$

Combining this relation with Eq. (45) leads to

$$\tilde{p}(\mathbf{r})=\tilde{\mathbf{a}}^T\boldsymbol{\alpha}(\mathbf{r})=\mathbf{a}^T\mathbf{A}_s[\mathbf{A}^H\mathbf{A}+\varepsilon\mathbf{I}]^{-1}\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})=\mathbf{a}^T\tilde{\boldsymbol{\alpha}}_s(\mathbf{r}), \tag{47}$$

meaning that we get the following expression for the relative error level in the sound pressure calculation:

$$E_p(\mathbf{r})\equiv 10\log\left(\frac{|\tilde{p}(\mathbf{r})-p(\mathbf{r})|^2}{|\mathbf{p}(\mathbf{r})|^2}\right)$$
$$=10\log\left(\frac{|\mathbf{a}^T[\tilde{\boldsymbol{\alpha}}_s(\mathbf{r})-\boldsymbol{\alpha}_s(\mathbf{r})]|^2}{|\mathbf{a}^T\boldsymbol{\alpha}_s(\mathbf{r})|^2}\right). \tag{48}$$

If in Eq. (48) we consider $\mathbf{a}$ as a vector of incoherent stochastic variables with equal rms values, and if we average the numerator and the dominator separately, then from Eq. (48) we obtain actually Eq. (43). For particular sound fields, however, $E_p(\mathbf{r})$ will differ from $E(\mathbf{r})$. In particular, $E_p(\mathbf{r})$ may be small because contributions from different wave-number components can cancel each other, which cannot happen for $E(\mathbf{r})$. $E(\mathbf{r})$ is therefore considered to be a better general measure of the error level, provided all components of the plane wave expansion coefficient vector $\mathbf{a}$ have almost equal amplitudes. This happens when the wave-number amplitude spectrum on the virtual source plane has a form close to a scaled version of $F(k_z)$ [see Eq. (16)].

E. Wave-number spectra of the reconstructed field

Wrap-around errors were entirely eliminated in plane wave SONAH by the use of continuous plane wave spectra, but spatial aliasing and wave-number domain leakage caused by the spatial sampling (discrete measurement points) and a finite measurement area cannot be avoided. For planar regular-grid arrays the spatial sampling introduces periodic replica in the wave-number domain, denoted as spatial aliasing, and the spatial window associated with the finite array size introduces leakage. This is basic Fourier acoustics theory.[3] But for irregular arrays, Fourier acoustic principles do not directly apply. In the description to be given here, the aliasing and leakage phenomena will be both covered by the term leakage.

To study wave-number leakage we need to calculate wave-number spectra associated with the sound field reconstructed by SONAH. Only free-field conditions and a planar measurement grid in the plane $z=0$ will be considered. The wave-number spectrum $\tilde{P}(k_x,k_y,z)$ of the reconstructed pressure in any plane parallel with the measurement plane is defined through the 2D Fourier transform:[3]

$$\tilde{P}(k_x,k_y,z)\equiv\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{p}(x,y,z)e^{j(k_x x+k_y y)}dxdy. \tag{49}$$

From Eqs. (49) and (15) we get

$$\tilde{P}(k_x,k_y,z)=\mathbf{q}^T\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{A}^H\boldsymbol{\alpha}(x,y,z)e^{j(k_x x+k_y y)}dxdy, \tag{50}$$

where the vector $\mathbf{q}$ is obtained from the vector $\mathbf{p}$ of measured pressure in the following way:

$$\mathbf{q}^T\equiv\mathbf{p}^T[\mathbf{A}^H\mathbf{A}+\varepsilon\mathbf{I}]^{-1}. \tag{51}$$

The integration in Eq. (50) can be worked out analytically using expression (21) for the elements of the vector $\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})$. If unit wave-number weighting is considered, $F(k_z)=1$, then the result is

$$\tilde{P}(k_x,k_y,z) = \frac{2\pi}{k^2} e^{-jk_z z} e^{j(k_z - k_z^*)z^+} \sum_i q_i e^{j(k_x x_i + k_y y_i)}, \quad (52)$$

q_i being the elements of $\mathbf{q}$ and $(x_i, y_i, 0)$ the coordinates of the measurement positions. Clearly, the plane wave spectrum depends on the z-coordinate only through the factor $e^{-jk_z z}$:

$$\tilde{P}(k_x,k_y,z) = \tilde{P}(k_x,k_y,0) e^{-jk_z z}, \quad (53)$$

which should be expected[3] since the reconstructed field satisfies the homogeneous wave equation in a homogeneous source-free half space.

An expression similar to Eq. (52) with $z=z^+$ can be derived by considering elements of the expansion coefficient vector $\tilde{\mathbf{a}}$ obtained from the equivalent EWM problem in Eq. (6). This agrees with the previously given statement that the vector $\tilde{\mathbf{a}}$ is a scaled wave-number spectrum in the virtual source plane. Use of Eqs. (6), (12), and (51) leads to the expression $\tilde{\mathbf{a}}^T = \mathbf{q}^T \mathbf{A}^H$, which agrees with Eq. (52) except for a constant scaling factor.

For leakage investigations we need to calculate the wave-number spectrum $\tilde{P}(k_x,k_y,z)$ resulting from a simulated measurement on an incident plane propagating or evanescent wave.

III. NUMERICAL INVESTIGATIONS OF RECONSTRUCTION ERRORS

A primary focus of the investigation in this section will be that of getting basic information about the regions of validity of the SONAH reconstructions for a couple of typical array geometries and as a function of frequency. Measurement errors will not be considered, and methods for automatic determination of the regularization parameter (GCV, L-curve, etc.) will not be investigated either. Instead the reader is referred to Gomes and Hansen.[16] The regularization parameter ε needed in Eqs. (11) and (27) will be determined from a specified dynamic range D (signal-to-noise-ratio) using the following formula from Steiner and Hald:[12]

$$\varepsilon = [\mathbf{A}^H \mathbf{A}]_{ii} 10^{-D/10}, \quad (54)$$

where $[\mathbf{A}^H \mathbf{A}]_{ii}$ is a diagonal element of the matrix $\mathbf{A}^H \mathbf{A}$. This relation was derived for a planar array and free-field conditions, in which case all diagonal elements are equal. Here we use, in general, the average value of the diagonal elements. Typical values of D lie in the range between 15 and 30 dB. Unless something else is specified, the following parameter choices have been used subsequently:

- dynamic range: $D=30$ dB
- identical virtual source planes for the test wave spectrum and for the wave spectrum used in SONAH prediction: $z_s^+ = z^+, z_s^- = z^-, \ldots$
- omni-directional weighting of the plane wave functions [see Eq. (19)]

A. Region of accurate pressure reconstruction in the measurement plane

We start with an investigation of the generic inherent error distributions in the array plane $z=0$ of two different irregular planar array designs. Both have a diameter equal to 0.75 m, and both are optimized for beamforming application up to at least 20 kHz. One is a spoke wheel array consisting of 13 identical six-element line arrays arranged as spokes in a wheel (see Fig. 5). The other is a sector wheel array[28] consisting of seven identical angular sectors with 12 microphones irregularly but uniformly distributed in each sector. The uniform distribution of the sector wheel array was maintained during the array geometry optimization in order to obtain good performance in connection with SONAH applications. Since the arrays are planar, we have to assume free-field conditions and use only a single virtual source plane in SONAH. An average element spacing is defined by calculating the average array area per element, and for both array geometries it is close to 7 cm. For a regular rectangular array, this spacing would support a frequency range up to around 2 kHz, and to avoid spatial aliasing up to that frequency, a standoff distance not smaller than the array grid spacing should be chosen. Since the virtual source plane should be positioned approximately one microphone spacing behind a real source surface at $z_0^+ = -7$ cm, we choose a virtual source plane 14 cm from the array plane, i.e., at $z^+ = -14$ cm.

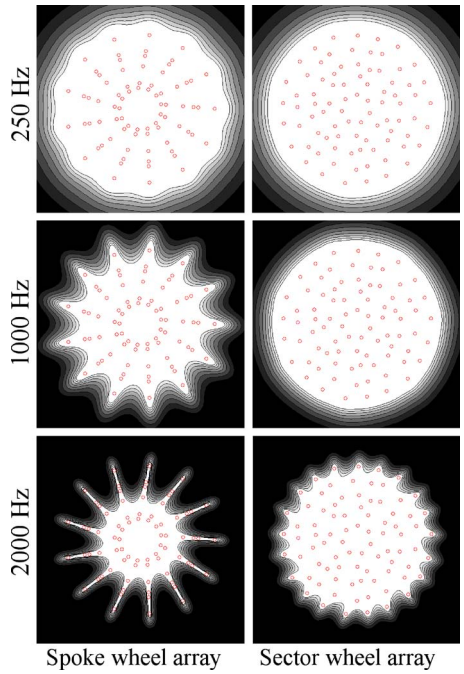


FIG. 5. (Color online) Inherent relative error level E of sound pressure reconstruction in the measurement plane for two irregular planar microphone arrays. Circles: measurement positions. White areas: $E<-20$ dB. Black areas: $E>-6$ dB. Gray-scale contour interval: 2 dB.

The contour plots in Fig. 5 show the inherent relative error level $E(\mathbf{r})$ of the sound pressure prediction in the array plane with the microphone positions indicated by circles. White areas represent inherent error levels $E(\mathbf{r})$ below -20 dB, and the gray contours show the degradation of the accuracy in steps of 2 dB to error levels larger than -6 dB represented by black. At low frequencies, both arrays can provide a good sound pressure representation across the entire array area, but already at 1 kHz the gaps in the microphone distribution of the spoke wheel array become too big to provide good representation in the gaps. The sector wheel array

provides acceptable pressure representation over the full array area all the way up to 2 kHz, which is clearly not the case for the spoke wheel array geometry. The spoke wheel array is very good for beamforming applications, though, and its support structure is easier to build in large sizes.

B. Region of accurate pressure reconstruction outside the measurement plane

We now turn to an investigation of the region of accurate sound pressure prediction in a plane normal to the array plane. Instead of the irregular arrays designed for beamforming, we consider regular 8×8 element single layer ($8\times8\times1$) and dual layer ($8\times8\times2$) arrays with 3 cm element spacing in the x- and y-directions, covering the area $0\leq x\leq21$ cm, $0\leq y\leq21$ cm. Both array types have a microphone layer in the xy-plane $z=0$, and the dual layer array has an additional identical layer at $z=3$ cm. We look at the relative sound pressure prediction error in the xz-plane $y=10.5$ cm, which is a symmetry plane of the arrays. Just like for the irregular planar arrays, we choose for the single layer regular array a virtual source plane at a distance from the array equal to two times the array element spacing, i.e., $z^+=-6$ cm. For the dual layer array we choose the same distance to the front virtual source plane and a distance to the rear virtual source plane equal to 19 times the element spacing, i.e., $z^+=-6$ cm, $z^-=60$ cm.

Figure 6 shows in the left column the inherent error level $E(\mathbf{r})$ of the single layer array (with a single virtual source plane) and in the right column the inherent error level $E(\mathbf{r})$ of the dual layer array (with two virtual source planes). The middle column shows the relative error $E_p(\mathbf{r})$ of the sound pressure prediction by the single layer array for a specific sound field: the field of a monopole point source at $(x,y,z)=(7.5,7.5,-6.0)$ cm. The mapping area extends 6 cm beyond the array area in both directions along the x-axis, and in the z-direction from $z=-3$ cm to a distance equal to four times the array element spacing in the positive z-direction. Clearly, the error distribution $E_p(\mathbf{r})$ of the particular sound field prediction is more irregular and, in general, lower than the inherent error. Significantly smaller errors are achieved in certain areas than predicted by the inherent error level. As explained in Sec. II D, this can be because error contributions from different plane wave components cancel each other in $E_p(\mathbf{r})$. The inherent error level avoids such cancellation effects and provides a much more general picture of the estimation accuracy of the method, although it does not give an upper bound on the error. For example, larger relative errors can occur in connection with very reactive sound fields, i.e., fields with a very high relative content of evanescent waves. The dual layer array is seen to extend significantly the region of accurate representation as compared to the single layer array, although the degrees of freedom in the sound field have also been increased by adding an extra virtual source plane.

The reason for the rather limited extent in the z-direction of the region with low prediction errors is that SONAH assumes nothing about the extent of the source distribution in the xy-directions. The field beyond some distance in front of the array cannot be accurately predicted because there can be contributions from source areas far from the array, for which the array measurement does not provide sufficient information. The minimum norm solution process then uses this lack of information about all the degrees of freedom in the field model to minimize the "energy" in the reconstructed field, resulting in the observed underestimation. One can think of the reconstructed field as being "focused" within the measurement region, representing well the measured sound pressure values. Before and after the measurement region, the field gradually defocuses. In order to provide, e.g., good far-field estimates, some assumption about the size of the source area would have to be integrated in the method. This could, for example, be done by the use of a SONAH formulation based on a limited distribution of monopole sources instead of plane propagating and evanescent waves. Brüel & Kjær SVM A/S has a patent pending, covering SONAH implementations based on distributions of monopoles, dipoles, and/or multipoles.

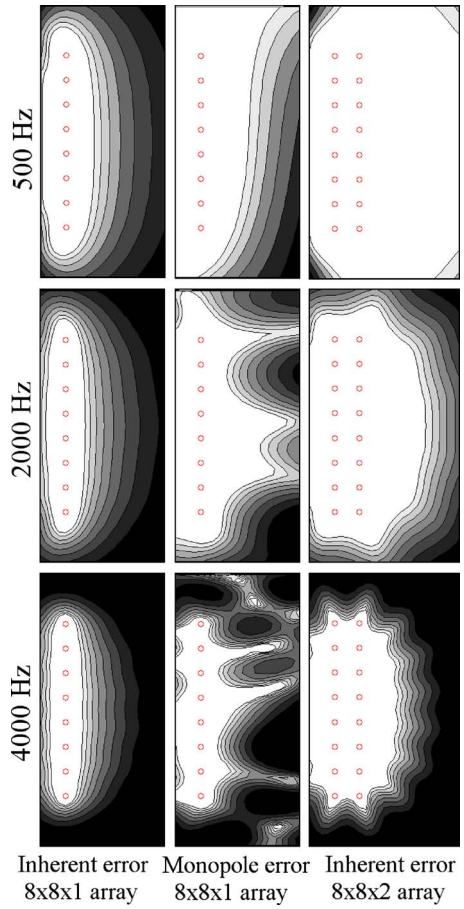


FIG. 6. (Color online) Relative error levels in the xz symmetry plane of single and dual layer 8×8 element microphone arrays. Left column: inherent error level E of the single layer array. Middle column: relative error of monopole pressure estimated using the single layer array. Right column: inherent error level E of the dual layer array.

C. Wave-number leakage

To illustrate some typical leakage behavior, which will be used to explain some observations in Sec. III D, we consider measurement with the single layer $8\times8\times1$ element

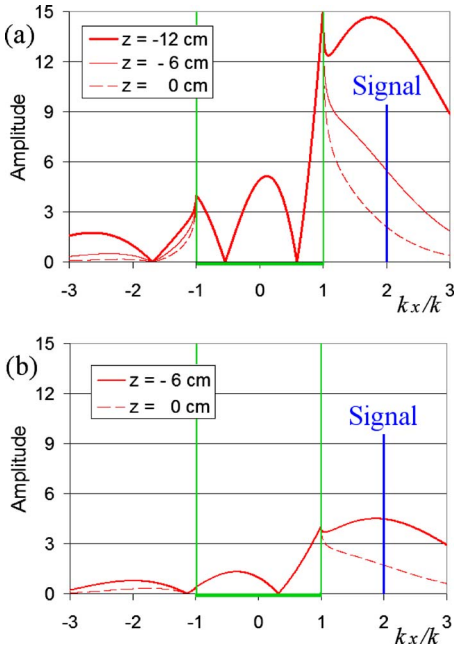


FIG. 7. (Color online) Wave-number (leakage) spectra at different z-coordinates for the same simulated measurement on a single evanescent wave with $k_x/k=2$, $k_y/k=0$. The two figures represent different virtual source plane distances: (a) $d=12$ cm and (b) $d=6$ cm. The frequency is 500 Hz, and the measurement plane is at $z=0$ cm.

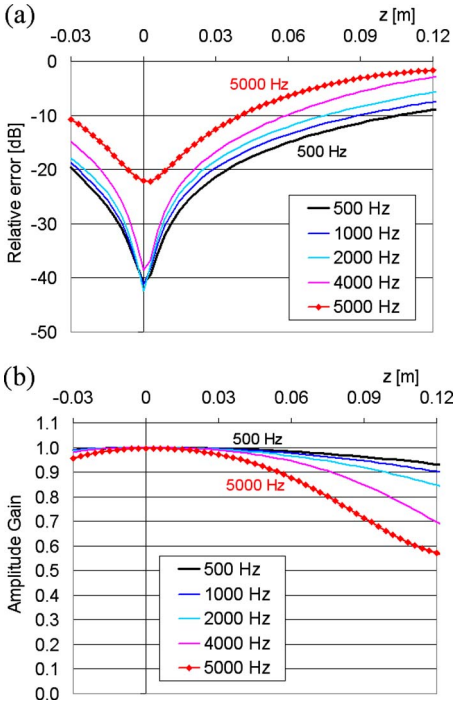


FIG. 8. (Color online) (a) Inherent relative error level E and (b) amplitude gain A on the symmetry axis normal to the measurement plane for an 8 $\times$8 element planar microphone array.

array of Sec. III B, and for simplicity a unit weighting ($F(k_z)=1$) of the plane wave functions is used. We consider a single 500 Hz evanescent wave propagating along the x-axis. The true wave-number spectrum has a single component at $(k_x,k_y)=(2k,0)$. Figure 7(a) shows slices along the k_x-axis of the wave-number spectra of Eq. (52) for a SONAH sound field reconstruction with virtual source plane distance $d=12$ cm. Three curves are shown, representing three different z-coordinates: $z=-d=-12$ cm, $z=-6$ cm, and $z=0$ cm. The horizontal axis shows normalized wave-numbers, k_x/k, so the points $k_x/k=\pm 1$ represent intersections with the radiation circle and the intervals $|k_x/k|>1$ represent evanescent waves. In the virtual source plane, $z=-d$, the wave-number spectrum has its maximum near the wave-number $k_x/k=2$ of the true wave field, and the leakage components decrease fairly smoothly away from that maximum. Figure 7(b) shows comparable SONAH wave-number spectra from the same simulated measurement, just with the virtual source plane at a shorter distance $d=6$ cm. Again, in the virtual source plane the maximum occurs near $k_x/k=2$, and the leakage spectrum decreases fairly smoothly away from that maximum. Thus, the leakage seems to be introduced in the virtual source plane, and the forward propagation from the virtual source plane to the measurement plane then strongly attenuates the evanescent wave components, while the plane wave components retain their amplitude. So in the measurement plane actually the plane wave leakage components are stronger than the component at the true wave-number of the original signal. If a larger value is chosen for the virtual source plane distance, then the reconstruction on the virtual source plane of a measured evanescent wave component requires a stronger amplification, leading to stronger leakage components that retain their amplitude inside the radiation circle at all z-coordinates. As mentioned previously, the EWM and SONAH reconstructions will always be very accurate at and near the measurement positions. The effect of the leakage described above is that a small portion of the evanescent waves is "converted" into propagating waves, which will erroneously contribute to the sound intensity and to the far field.

D. Sensitivity of reconstruction accuracy to changing parameters of SONAH

To limit the amount of data in the investigation of sensitivity to parameter changes, we look at the prediction errors only along the center axis normal to the array. Again, only the single layer $8\times8\times1$ array of Sec. III B will be considered. Figure 8(a) shows the inherent error level for the same condition as that represented in the left column of Fig. 6, so what we see is just a set of horizontal slices through the middle of the contour plots in the left column of Fig. 6. Two additional frequencies have been included in Fig. 8(a): 1 kHz and the higher limiting frequency 5 kHz of the 3 cm spaced regular array grid. Clearly, at 5 kHz the inherent error level has started a quick increase, even in the measurement plane, but an acceptable reconstruction can still be achieved. Figure 8(b) shows the corresponding curves of amplitude gain $A(\mathbf{r})$. Even at 5 kHz the amplitude is still accurately predicted within one array grid spacing from the array plane, but outside that interval an increasing underestimation is observed. As mentioned previously, this can be seen as a consequence of the fact that the equivalent EWM problem will perform a least norm estimation of the plane wave spectrum, thus minimizing the energy in the reconstructed sound field.

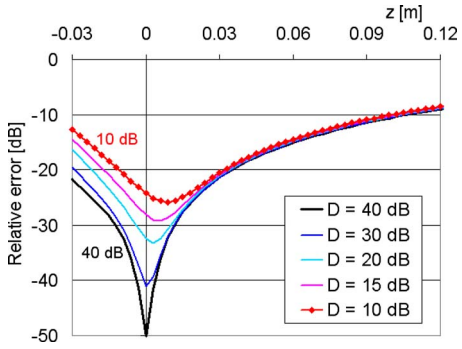


FIG. 9. (Color online) Inherent relative error level E on the symmetry axis normal to the array plane for an 8×8 element planar microphone array at 500 Hz. The curves represent different dynamic ranges, D, used in the SONAH processing.

In Fig. 8 the regularization parameter was determined from Eq. (54) with a fixed dynamic range D equal to 30 dB. The next step is to see how the reconstruction error changes when the dynamic range D is changed. Figure 9 presents the sensitivity of the inherent error level E to changes in D at a low frequency, 500 Hz. A set of curves is labeled with the corresponding values of D, which range from 10 to 40 dB. The prediction of pressure further away from the source region than the measurement plane is insensitive, while the prediction closer to the source region improves with increasing values of D. This should be expected at low frequencies since a smaller value of D will limit the reconstruction of the evanescent waves and since the evanescent waves constitute a bigger part of the sound field at low frequencies and close to the source region. Figure 10 presents corresponding curves for the inherent amplitude gain A in the SONAH estimation at a high frequency, 4000 Hz. Here, an increasing underestimation is observed at all positions (even further away from the source region) when the value of D becomes smaller than around 20 dB. The explanation must be that at high frequencies the propagating plane waves constitute a bigger part of the plane wave spectrum and thus of the singular value spectrum of the SONAH auto-correlation matrix $\mathbf{A}^H\mathbf{A}$. When D is reduced beyond some point, then even parts of the propagating wave components will be significantly attenuated by the regularization.

The question to be investigated now is how the error

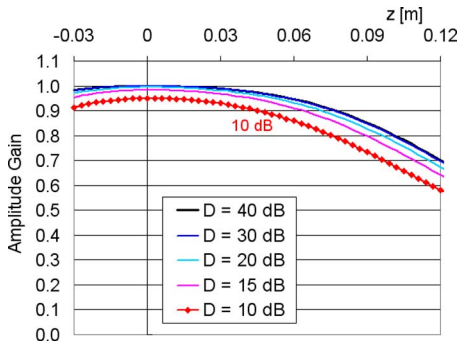


FIG. 10. (Color online) Amplitude gain A on the symmetry axis normal to the array plane for an 8×8 element planar microphone array at 4000 Hz. The curves represent different dynamic ranges, D, used in the SONAH processing.

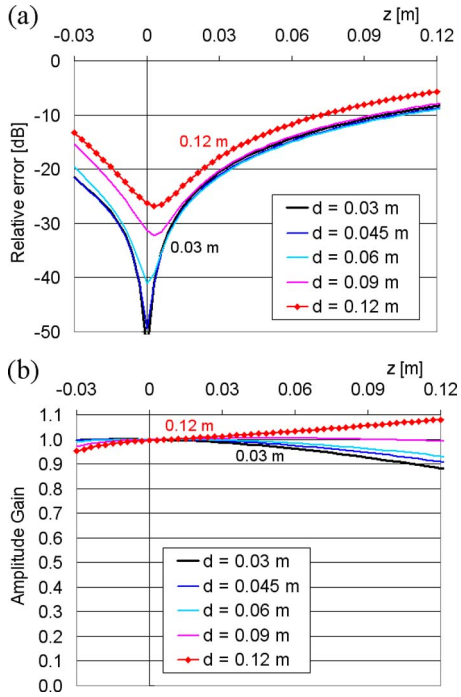


FIG. 11. (Color online) (a) Inherent relative error level E and (b) amplitude gain A on the symmetry axis normal to the array plane for an 8×8 element planar microphone array at 500 Hz. The curves represent different virtual source plane distances, d, used in the SONAH processing, while the virtual source distance defining the actual content of evanescent waves is kept constant at 6 cm.

level changes when the virtual source planes ($z=z_s^+,\ldots$) used for the test wave spectra differ from the virtual source planes ($z=z^+,\ldots$) used in the SONAH algorithm. This can, to some extent, show the sensitivity of SONAH to the wave-number amplitude spectrum of a real sound field having a different shape from the one for which the SONAH prediction weight vector $\mathbf{c}$ was calculated [see Eqs. (18), (16), and (10)]. Only free-field conditions and the use of a single virtual source plane shall be considered. For the results presented in Figs. 11 and 12, we have retained the previously used virtual source plane $z=z_s^+=-6$ cm for the test wave spectrum, i.e., in the calculation of $\mathbf{A}_s^H\mathbf{A}_s$, $\mathbf{A}_s^H\boldsymbol{\alpha}_s$, and $\boldsymbol{\alpha}_s^H\boldsymbol{\alpha}_s$ used in Eqs. (39), (40), and (42), while a varying virtual source plane $z=z^+=-d$ at distance d from the array has been used for the SONAH predictions. A large value of d will put a low weight on the evanescent waves in the SONAH estimation. Without regularization the smallest generic inherent error should be achieved with $z^+=z_s^+$ i.e., $d=6$ cm: The SONAH prediction weight vector $\mathbf{c}$ minimizes the residual in Eq. (10), and with $z^+=z_s^+$ that will automatically minimize the generic inherent error for the test field. But because of the regularization ($D=30$ dB), the inherent error may be slightly smaller with $z^+\neq z_s^+$. Figure 11 presents the sensitivity of the inherent error level E and the amplitude gain A to changes in d at a low frequency, 500 Hz, and Fig. 12 shows the corresponding results for a high frequency, 4000 Hz. Each plot contains a set of curves labeled with the corresponding values of d, which range from 3 to 12 cm.

At 500 Hz the error level is rather insensitive to a too small virtual source distance, d, i.e., a too high weight on the

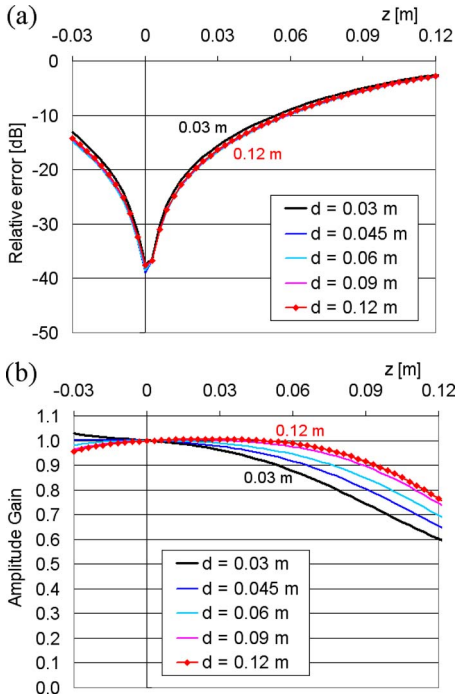


FIG. 12. (Color online) (a) Inherent relative error level E and (b) amplitude gain A on the symmetry axis normal to the array plane for an 8×8 element planar microphone array at 4000 Hz. The curves represent different virtual source plane distances, d, used in the SONAH processing, while the virtual source distance defining the actual content of evanescent waves is kept constant at 6 cm.

evanescent waves in the reconstruction, but a too large distance eventually results in a slight overestimation ($A>1$) of the level further away from the source region than the measurement plane. The explanation is the one outlined in Sec. III C above: Evanescent wave components leak into propagating components, leading to overestimation of the sound pressure further away from the source region and underestimation closer to the source region. As explained in Sec. III C, this leakage increases with increasing virtual source plane distance, d. Looking at the more general increase in the inherent error level for $z<0$, occurring also when d is large, there is an additional source of error: Because too low weight is put on the evanescent waves in the SONAH calculation matrices (due to the large virtual source distance), the chosen regularization parameter will prevent some of these wave components from being sufficiently reconstructed.

At 4 kHz the evanescent waves constitute, in general, a much smaller part of the sound field, so the sensitivity to putting more or less weight on them in the SONAH calculation is smaller. The biggest effect is on the amplitude gain when a too small virtual source distance, d, is used, putting a higher weight on the evanescent waves in the reconstruction than suggested by the actual wave-number spectrum. Close to the source region an overestimation is observed, while further away a heavier underestimation than normal is seen. The explanation is probably leakage, just the other way than described in Sec. III C: Some small propagating wave components will leak into evanescent waves, which will be attenuated away from the source area and amplified toward the source area.

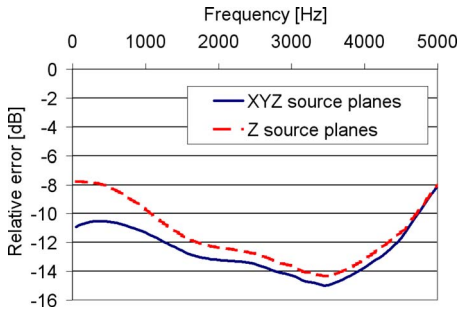


FIG. 13. (Color online) Accuracy improvement gained by adding plane wave functions from all six sides instead of from just two sides in a SONAH calculation of 3D vector intensity from a single monopole point source in the center plane of a dual layer 8×8 element microphone array. The point source is 6 cm outside of the array area.

E. Sources in the array plane: Need for six virtual source planes

Figure 13 shows an example of the accuracy improvement gained by adding plane wave functions from all six sides instead of from just two sides, allowing a correct modeling of evanescent wave components that decay along directions in the xy-plane. A simulated measurement is performed with the same $8\times8\times2$ dual layer array geometry as the one described in Sec. III B. A single monopole point source is placed at $(x,y,z)=(-6.0,7.5,1.5)$ cm, i.e., in the center plane between the two array layers and 6 cm outside the array area. Clearly, this monopole will create evanescent wave types that decay along directions in the xy-plane. SONAH is used to predict the intensity vectors at the measurement positions in the front array plane $z=0$ cm, based on the 128 measured pressure values. The relative average error level is then calculated using the formula

$$L_{\mathrm{err}}=10\log_{10}\left(\frac{\sum|\mathbf{I}_i-\tilde{\mathbf{I}}_i|}{\sum|\mathbf{I}_i|}\right), \tag{55}$$

where $\mathbf{I}_i$ is the true intensity vector at calculation point number i, $\tilde{\mathbf{I}}_i$ is the SONAH estimate of the same vector, and the summations are over i. The SONAH calculation is performed using a 30 dB dynamic range, $D=30$ dB, and the virtual source planes are chosen to be 6 cm away from the array in all directions: $z^-=9$ cm, $z^+=-6$ cm, $x^-=27$ cm, $x^+=-6$ cm, $y^-=27$ cm, and $y^+=-6$ cm. Thus, the monopole is located on the boundary of the box defined by the six virtual source planes. Looking at Fig. 13, the low-frequency intensity estimation is seen to be significantly more accurate when all six source planes are present than when only those parallel with the xy-plane are present. The improvement vanishes, however, with increasing frequency because the evanescent waves constitute a smaller part of the sound field at higher frequencies.

IV. CONCLUSIONS

The basic theory and a revised mathematical formulation of SONAH have been given, first in a general form allowing arbitrary expansion functions and then for the special cases of one, two, or six sets of plane wave functions representing

different virtual source planes. The relation to the EWM method was outlined and used in the explanation of some important properties of the SONAH method, in particular the underestimation beyond some distance from the measurement region. To obtain a general measure of the prediction accuracy of SONAH, a set of formulas for the inherent error level was derived. These have proven to be very useful to visualize and determine the regions of validity of the SONAH predictions with given array geometries. Two typical types of array geometry were treated in the paper: first, planar irregular arrays optimized for broadband beamforming application and second, small single and dual layer regular array grids. The sensitivity of the inherent error level to changes in parameters of the SONAH algorithm was then investigated and shown to be fairly small for parameter values within realistic intervals. This investigation covered the choice of regularization parameter and the setting of the virtual source plane distance, which controls the weight applied to the evanescent waves in the SONAH estimation. Finally, a simulated measurement with a monopole sound source in the array plane was used to check the benefit of adding six virtual source planes instead of just two: Some improvement was achieved (up to 3 dB at low frequencies) due to the added ability to correctly model evanescent waves that decay along directions in the array plane.

APPENDIX A: REGULARIZED LEAST-SQUARES AND LEAST NORM SOLUTIONS

The regularized, generalized inverses $\mathbf{A}^{+}_{LS,\varepsilon}$ and $\mathbf{A}^{+}_{LN,\varepsilon}$ of an arbitrary matrix $\mathbf{A}$ associated with the least-squares (5) and least norm (6) solution formulas, respectively, are

$$\mathbf{A}^{+}_{LS,\varepsilon} \equiv (\mathbf{A}^H\mathbf{A} + \varepsilon\mathbf{I})^{-1}\mathbf{A}^H, \tag{A1}$$

$$\mathbf{A}^{+}_{LN,\varepsilon} \equiv \mathbf{A}^H(\mathbf{A}\mathbf{A}^H + \varepsilon\mathbf{I})^{-1}, \tag{A2}$$

where the regularization parameter ε is positive real. Premultiplication with the non-singular square matrix $(\mathbf{A}^H\mathbf{A}+\varepsilon\mathbf{I})$ and post-multiplication with the non-singular square matrix $(\mathbf{A}\mathbf{A}^H+\varepsilon\mathbf{I})$ on both $\mathbf{A}^{+}_{LS,\varepsilon}$ and $\mathbf{A}^{+}_{LN,\varepsilon}$ show that $\mathbf{A}^{+}_{LS,\varepsilon}=\mathbf{A}^{+}_{LN,\varepsilon}$:

$$\begin{aligned}(\mathbf{A}^H\mathbf{A} + \varepsilon\mathbf{I})\mathbf{A}^{+}_{LS,\varepsilon}(\mathbf{A}\mathbf{A}^H + \varepsilon\mathbf{I}) &= \mathbf{A}^H(\mathbf{A}\mathbf{A}^H + \varepsilon\mathbf{I}) \\ &= \mathbf{A}^H\mathbf{A}\mathbf{A}^H + \varepsilon\mathbf{A}^H, \\ (\mathbf{A}^H\mathbf{A} + \varepsilon\mathbf{I})\mathbf{A}^{+}_{LN,\varepsilon}(\mathbf{A}\mathbf{A}^H + \varepsilon\mathbf{I}) &= (\mathbf{A}^H\mathbf{A} + \varepsilon\mathbf{I})\mathbf{A}^H \\ &= \mathbf{A}^H\mathbf{A}\mathbf{A}^H + \varepsilon\mathbf{A}^H. \end{aligned} \tag{A3}$$

APPENDIX B: REDUCTION IN SONAH CORRELATION INTEGRALS

We consider the integral expressions for the correlation functions $[\mathbf{A}^H\mathbf{A}]_{ij}$ and $[\mathbf{A}^H\boldsymbol{\alpha}(\mathbf{r})]_i$ in Eqs. (20) and (21). Since they have exactly the same form, we consider only the first of them:

$$[\mathbf{A}^H\mathbf{A}]_{ij} = \frac{1}{2\pi k^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi^*_{\mathbf{k}}(\mathbf{r}_i)\Phi_{\mathbf{k}}(\mathbf{r}_j)dk_xdk_y, \tag{B1}$$

which through application of Eqs. (16) and (19) becomes

$$\begin{aligned}[\mathbf{A}^H\mathbf{A}]_{ij} = \frac{F_0^2}{2\pi k^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{k}{|k_z|} \\ \times e^{j[k_x(x_i-x_j)+k_y(y_i-y_j)]}e^{j[k_z^*z_i-k_zz_j+(k_z-k_z^*)z^+]}dk_xdk_y.\end{aligned} \tag{B2}$$

To reduce the integral we transform to polar coordinates in two dimensions in both the wave-number domain and in the spatial domain:

$$\begin{pmatrix}k_x\\k_y\end{pmatrix} = K\begin{pmatrix}\cos(\psi)\\\sin(\psi)\end{pmatrix}, \quad \begin{pmatrix}x_i-x_j\\y_i-y_j\end{pmatrix} = R_{ij}\begin{pmatrix}\cos(\varphi_{ij})\\\sin(\varphi_{ij})\end{pmatrix}. \tag{B3}$$

With these definitions, the argument of the first exponential function in Eq. (B2) can be rewritten as

$$k_x(x_i-x_j) + k_y(y_i-y_j) = KR_{ij}\cos(\psi-\varphi_{ij}), \tag{B4}$$

and since k_z does not depend on the polar angle ψ, the polar angle integration can be performed analytically with the following result:[29]

$$\begin{aligned}[\mathbf{A}^H\mathbf{A}]_{ij} &= \frac{F_0^2}{2\pi k^2}\int_0^{\infty}\int_0^{2\pi}\frac{k}{|k_z|} \\ &\quad\times e^{jKR_{ij}\cos(\psi-\varphi_{ij})}e^{j[k_z^*z_i-k_zz_j+(k_z-k_z^*)z^+]}Kd\psi dK \\ &= \frac{F_0^2}{k^2}\int_0^{\infty}\frac{k}{|k_z|}J_0(KR_{ij})e^{j[k_z^*z_i-k_zz_j+(k_z-k_z^*)z^+]}KdK,\end{aligned} \tag{B5}$$

where J_0 is the Bessel function of order zero. Equations (17) and (B3) show that k_z can be expressed as

$$k_z = \begin{cases}\sqrt{k^2-K^2}, & K\le k\\ -j\sqrt{K^2-k^2}, & K>k,\end{cases} \tag{B6}$$

which implies that $k_z^*=k_z$ for $K\le k$ and $k_z^*=-k_z$ for $K>k$. Use of this in Eq. (B5) leads to

$$\begin{aligned}[\mathbf{A}^H\mathbf{A}]_{ij} &= \frac{F_0^2}{k^2}\int_0^{k}\frac{k}{|k_z|}J_0(KR_{ij})e^{jk_z(z_i-z_j)}KdK \\ &\quad+\frac{F_0^2}{k^2}\int_k^{\infty}\frac{k}{|k_z|}J_0(KR_{ij})e^{-jk_z(z_i+z_j-2z^+)}KdK,\end{aligned} \tag{B7}$$

where we can insert Eq. (B6) to obtain

$$\begin{aligned}[\mathbf{A}^H\mathbf{A}]_{ij} &= \frac{F_0^2}{k^2}\int_0^{k}\frac{k}{\sqrt{k^2-K^2}}J_0(KR_{ij})e^{j\sqrt{k^2-K^2}(z_i-z_j)}KdK \\ &\quad+\frac{F_0^2}{k^2}\int_k^{\infty}\frac{k}{\sqrt{K^2-k^2}}J_0(KR_{ij})e^{-\sqrt{K^2-k^2}(z_i+z_j-2z^+)}KdK.\end{aligned} \tag{B8}$$

Here, the square roots are a problem in connection with numerical integration. To get rid of them, we introduce the following variable substitutions for K:

TABLE I. Integral expressions for the functions used to set up the SONAH correlation matrices. Both unit weighting $[F(k_z)=1]$ and omni-directional weighting $[F(k_z)=\sqrt{k/|k_z|}]$ of the plane wave functions are covered.

	$F(k_z)$	
	1	$\sqrt{k/\lvert k_z\rvert}$
$I_p(\rho,\zeta)$	$\int_0^{\pi/2} J_0(\rho\sin(\theta))e^{j\zeta\cos(\theta)}\sin(\theta)\cos(\theta)d\theta$	$\int_0^{\pi/2} J_0(\rho\sin(\theta))e^{j\zeta\cos(\theta)}\sin(\theta)d\theta$
$II_p(\rho,\zeta)$	$\int_0^{\infty} J_0(\rho\sqrt{1+v^2})e^{-\zeta v}v\,dv$	$\int_0^{\infty} J_0(\rho\sqrt{1+v^2})e^{-\zeta v}dv$
$I_z(\rho,\zeta)$	$\int_0^{\pi/2} J_0(\rho\sin(\theta))e^{j\zeta\cos(\theta)}\sin(\theta)\cos^2(\theta)d\theta$	$\int_0^{\pi/2} J_0(\rho\sin(\theta))e^{j\zeta\cos(\theta)}\sin(\theta)\cos(\theta)d\theta$
$II_z(\rho,\zeta)$	$\int_0^{\infty} J_0(\rho\sqrt{1+v^2})e^{-\zeta v}v^2dv$	$\int_0^{\infty} J_0(\rho\sqrt{1+v^2})e^{-\zeta v}v\,dv$
$I_{xy}(\rho,\zeta)$	$\int_0^{\pi/2} \frac{J_1(\rho\sin(\theta))}{\rho\sin(\theta)} e^{j\zeta\cos(\theta)}\sin^3(\theta)\cos(\theta)d\theta$	$\int_0^{\pi/2} \frac{J_1(\rho\sin(\theta))}{\rho\sin(\theta)} e^{j\zeta\cos(\theta)}\sin^3(\theta)d\theta$
$II_{xy}(\rho,\zeta)$	$\int_0^{\infty} \frac{J_1(\rho\sqrt{1+v^2})}{\rho\sqrt{1+v^2}} e^{-\zeta v}(1+v^2)v\,dv$	$\int_0^{\infty} \frac{J_1(\rho\sqrt{1+v^2})}{\rho\sqrt{1+v^2}} e^{-\zeta v}(1+v^2)dv$

$$K=\begin{cases} k\sin(\theta), & K\le k \\ k\sqrt{1+v^2}, & K>k, \end{cases} \tag{B9}$$

with new variables θ and v, and as a result Eq. (B8) becomes

$$[\mathbf{A}^H\mathbf{A}]_{ij}=F_0^2\int_0^{\pi/2} J_0(kR_{ij}\sin(\theta))e^{jk(z_i-z_j)\cos(\theta)}\sin(\theta)d\theta + F_0^2\int_0^{\infty} J_0(kR_{ij}\sqrt{1+v^2})e^{-k(z_i+z_j-2z^+)v}dv, \tag{B10}$$

where the functions to be integrated are smooth and well suited for numerical integration.

APPENDIX C: 2D FUNCTIONS USED TO SET UP SONAH CORRELATION MATRICES

In Table I, J_0 and J_1 are the Bessel functions of order 0 and 1, respectively. The author has been able to find partial analytic expressions only for a couple of the integrals in the table and for some special parameter values, so the integrals in general have to be calculated by numerical integration. The number of integrations can be reduced very much by generating (by numerical integration) and storing 2D tables for discrete values of the variables ρ and ζ. To calculate the integrals for specific values of ρ and ζ, interpolations have to be performed between the values in the tables. Alternatively, 2D approximation functions covering different domains can be constructed. Brüel & Kjaer SVM A/S has a patent pending, covering the use of tables and approximation functions to speed up the execution of SONAH implementations.

The 2D functions of Table I are, in practice, used only through the following combination functions:

$$III_p(\rho,\zeta_1,\zeta_2)\equiv I_p(\rho,\zeta_1)+II_p(\rho,\zeta_2), \tag{C1}$$

$$III_z(\rho,\zeta_1,\zeta_2)\equiv I_z(\rho,\zeta_1)-jII_z(\rho,\zeta_2), \tag{C2}$$

$$III_{xy}(\rho,\zeta_1,\zeta_2)\equiv I_{xy}(\rho,\zeta_1)+II_{xy}(\rho,\zeta_2). \tag{C3}$$

[1]E. G. Williams, J. D. Maynard, and E. J. Skudrzyk, "Sound source reconstruction using a microphone array," J. Acoust. Soc. Am. **68**, 340–344 (1980).
[2]J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[3]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, 1999).
[4]J. Hald, "Reduction of spatial windowing effects in acoustical holography," Proceedings of Inter-Noise (1994).
[5]K. Saijyou and S. Yoshikawa, "Reduction methods of the reconstruction error for large-scale implementation of near-field acoustical holography," J. Acoust. Soc. Am. **110**, 2007–2023 (2001).
[6]E. G. Williams, "Continuation of acoustic near-fields," J. Acoust. Soc. Am. **113**, 1273–1281 (2003).
[7]E. G. Williams and B. H. Houston, "Fast Fourier transform and singular value decomposition formulations for patch nearfield acoustical holography," J. Acoust. Soc. Am. **114**, 1322–1333 (2003).
[8]K. Saijyou and H. Uchida, "Data extrapolation method for boundary element method-based near-field acoustical holography," J. Acoust. Soc. Am. **115**, 785–796 (2004).
[9]M. Lee and J. S. Bolton, "Patch near-field acoustical holography in cylindrical geometry," J. Acoust. Soc. Am. **118**, 3721–3732 (2005).
[10]Z. Wang and S. F. Wu, "Helmholtz equation-least-squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[11]S. F. Wu, "On reconstruction of acoustic fields using the Helmholtz equation-least-squares method," J. Acoust. Soc. Am. **107**, 2511–2522 (2000).
[12]R. Steiner and J. Hald, "Near-field acoustical holography without the errors and limitations caused by the use of spatial DFT," Int. J. Acoust. Vib. **6**, 83–89 (2001).
[13]J. Hald, "Patch near-field acoustical holography using a new statistically optimal method," Proceedings of Inter-Noise (2003).
[14]C.-T. Cho, J. S. Bolton, and J. Hald, "Source visualization by using statistically optimized near-field acoustical holography in cylindrical coordinates," J. Acoust. Soc. Am. **118**, 2355–2364 (2005).
[15]J. Hald, "Patch holography in cabin environments using a two-layer handheld array and an extended SONAH algorithm," Proceedings of Euronoise (2006).
[16]J. Gomes and P. C. Hansen, "A study on regularization parameter choice in near-field acoustical holography," Proceedings of Euronoise (Acoustics'08) (2008).
[17]J. Hald and J. Gomes, "A comparison of two patch NAH methods," Proceedings of Inter-Noise (2006).
[18]F. Jacobsen and V. Jaud, "Statistically optimized near-field acoustic holography using an array of pressure-velocity probes," J. Acoust. Soc. Am. **121**, 1550–1558 (2007).
[19]J. Hald, J. Mørkholt, P. Hardy, D. Trentin, M. Bach-Andersen, and G. Keith, "Array based measurement of radiated and absorbed sound intensity components," Proceedings of Euronoise (Acoustics'08) (2008).

[20]A. Sarkissian, "Method of superposition applied to patch near-field acoustical holography," J. Acoust. Soc. Am. **118**, 671–678 (2005).
[21]N. Valdivia and E. G. Williams, "Study on the comparison of the methods of equivalent sources and boundary element methods for near-field acoustic holography," J. Acoust. Soc. Am. **120**, 3694–3705 (2006).
[22]J. Gomes, "Patch holography using a double layer microphone array," Proceedings of Inter-Noise (2007).
[23]N. Valdivia and E. G. Williams, "Approximations of inverse boundary element methods with partial measurements of the pressure field," J. Acoust. Soc. Am. **123**, 109–120 (2008).
[24]N. Valdivia and E. G. Williams, "Implicit methods of solution to integral formulations in boundary element method based nearfield acoustic holography," J. Acoust. Soc. Am. **116**, 1559–1572 (2004).
[25]S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, England, 2004).
[26]P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (Society for Industrial and Applied Mathematics, Philadelphia, 1997).
[27]J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1996).
[28]J. Hald, "Array designs optimized for both low-frequency NAH and high-frequency beamforming," Proceedings of Inter-Noise (2004).
[29]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).

Multiple-point statistical room correction for audio reproduction: Minimum mean squared error correction filtering

Fredrik Lingvall[a] and Lars-Johan Brännmark[b]
Signals and Systems Group, Department of Engineering Sciences, Uppsala University, P.O. Box 528, SE-751 20 Uppsala, Sweden

(Received 17 October 2008; accepted 30 December 2008)

This paper treats the problem of correction of loudspeaker and room responses using a single source. The objective is to obtain a linear correction filter, which is robust with respect to listener movement within a predefined region-of-interest. The correction filter is based on estimated impulse responses, obtained at several positions, and a linear minimum mean squared error criteria. The impulse responses are estimated using a Bayesian approach that takes both model errors and measurement noise into account, which results in reliable impulse response estimates and a measure of the estimation errors. The correction filter is then constructed by using information from both the estimated impulse response coefficients and their associated estimation errors. Furthermore, in the optimization criteria a time-dependent reflection filter is introduced, which attenuates the high frequency parts of the reflected responses, that is, the parts of the responses that cannot be compensated with a single source system. The resulting correction filter is shown to significantly improve both the temporal and spectral properties of the responses compared to the uncorrected system, and, furthermore, the obtained correction filter has a low level of pre-ringing.
 [DOI: 10.1121/1.3075615]

PACS number(s): 43.60.Uv, 43.55.Br, 43.60.Pt [EJS] Pages: 2121–2128

I. INTRODUCTION

The aim of digital room correction is, in general, to improve audio reproduction in, for example, domestic or automotive sound systems. Improve means here to compensate, or at least reduce, the distortions of the signal due to both a non-perfect source and acoustic diffraction and room effects. In digital room correction this is normally accomplished by first acquiring data to estimate the room impulse responses and then construct a correction filter based on the estimated impulse responses.[1]

It is known that direct inversion of acoustic impulse responses acquired at a single listening point results in excessive boost at some frequencies due to interference induced deep spectral notches in the impulse responses, which results in high noise amplification,[2,3] and, furthermore, the correction performance quickly deteriorates when the listener moves away from the reference correction point.[4,5] In this area of application, spatial robustness is essential since the position of the listener cannot be expected to be fixed. In the literature, several approaches to robust design of correction filters have been studied. Examples of such approaches include complex smoothing,[6] multiple-point equalization of common poles,[7] derivative constrained equalization,[8] spatial averaging with fuzzy clustering of room responses,[9] time-domain inversion with frequency-dependent regularization,[2] and robust multiple-point designs.[4,10,11]

Another important issue in correction filter design is to try to avoid non-causal preresponses of the correction filter; non-causal means here the filter effects that occur before the main correction filter delay. Allowing a delay in the system normally improves the correction performance,[12] but it also implies a mixed phase character of the correction filters. It is known that mixed phase designs generate a pre-response in the filter, which may generate an audible pre-ringing in the corrected system.[3,13]

This paper is concerned with compensation of loudspeaker and room effects, using a single source, by means of deconvolution of acoustic impulse responses based on data measured at more than one listening point. The objective is to obtain correction filters, with a low amount of pre-ringing, that are robust in the sense that the position of the listener is allowed to vary within a predefined region-of-interest while still obtaining a good, or at least reasonable, compensation for room and loudspeaker characteristics. The reason for choosing such an objective is that we must realize that the one cannot perfectly reproduce the input signal at every point in a region-of-interest. This is mainly for two reasons: (i) we are using a single source so we cannot control (beamform) the sound field spatially and (ii) we do not know the impulse responses at every point in advance and we must, therefore, estimate them based on data acquired at a finite set of observation points. Hence, we only have knowledge of the characteristics of the sound reproduction system at some points, and, furthermore, the estimated impulse responses have a finite length and are based on noisy measurements resulting in limited accuracy of the impulse response estimates.

In this paper we will discuss a statistical inferential[14] method for estimating acoustic impulse responses as well as a method to design robust compensation filters based on (i)

[a]Electronic mail: fredrik.lingvall@gmail.com
[b]Electronic mail: lars-johan.brannmark@angstrom.uu.se

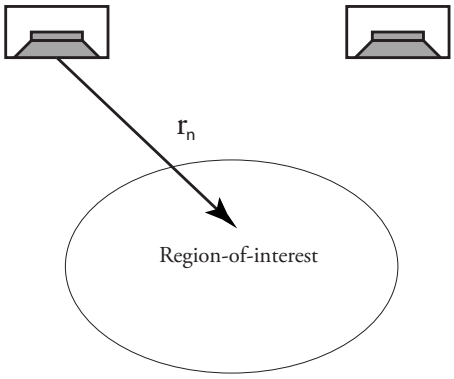


FIG. 1. Typical measurement setup.

impulse responses estimated from data acquired at several observation points and (ii) a linear minimum mean squared criteria. Note that we are seeking a linear finite impulse response filter that can be used for real time compensation.

This paper is organized as follows: in Sec. II a linear matrix-based impulse response model is introduced, which then is used in the estimation of the acoustic impulse responses, which is discussed in Sec. III. Furthermore, In Sec. IV the robust multiple-point design method for correction filters are described as well as how room reflections can be treated by means of a non-stationary filtering approach. In Sec. V the experiments are discussed, and, finally, in Sec. VI a discussion and the conclusions are given.

II. A LINEAR IMPULSE RESPONSE MODEL

Consider the loudspeaker setup shown in Fig. 1. We measure the acoustic pressure wave form at an observation point $\mathbf{r}_n=(x_n,y_n,z_n)$ using a microphone. If we let the loudspeaker be located at origo, $\mathbf{r}=(0,0,0)$, then we can express the received signal at the nth position as

$$y(\mathbf{r}_n,t)=h(\mathbf{r}_n,t)*u(t)+e(t), \tag{1}$$

where $*$ denotes temporal convolution. Note that the impulse response $h(\mathbf{r}_n,t)$ is a function of position due to diffraction effects of the loudspeaker and room effects, such as reflections and room modes; the impulse response, $h(\mathbf{r}_n,t)$, is comprised of several parts, such as the amplifier response, the electro-acoustical response of the loudspeaker, the acoustic (diffraction) response of the loudspeaker, the room impulse response, and the response of the microphone.

Here we are concerned with digital processing of the acoustic data; hence we need a discrete model of the system. This can readily be obtained by sampling the impulse responses, input signals, and the acquired data; here we represent, by convention, the data $\mathbf{y}_n$, the input signal $\mathbf{u}$, and the impulse responses $\mathbf{h}_n$ by column vectors. The discrete form of Eq. (1) can then be expressed as

$$\mathbf{y}_n=\mathbf{h}_n*\mathbf{u}+\mathbf{e}, \tag{2}$$

where $*$ now denotes discrete temporal convolution. Furthermore, a discrete-time convolution of two vectors $\mathbf{u}$ and $\mathbf{h}$ can be expressed as a matrix-vector multiplication of the form

$$\mathbf{h}*\mathbf{u}=\mathbf{H}\mathbf{u}=\mathbf{u}*\mathbf{h}=\mathbf{U}\mathbf{h}, \tag{3}$$

where the matrices $\mathbf{H}$ and $\mathbf{U}$ have a Toeplitz form consisting of delayed versions of $\mathbf{h}_n$ and $\mathbf{u}$, respectively.[12,15] If $\mathbf{h}_n$ is of length L_h and $\mathbf{u}$ is of length L_u, then the length of the discrete-time convolution, $\mathbf{h}_n*\mathbf{u}$, will be L_h+L_u-1. We will assume, for the remaining part of the paper, that when a matrix convolution, of the form (3), is used then the Toeplitz matrices involved have a dimension such that the length of the resulting vector is the same as for the normal form of the discrete-time convolution. Now, by using the matrix-vector notation in Eq. (3), the model (2) becomes

$$\mathbf{y}_n=\mathbf{H}_n\mathbf{u}+\mathbf{e}=\mathbf{U}\mathbf{h}_n+\mathbf{e}. \tag{4}$$

III. ESTIMATING ACOUSTIC IMPULSE RESPONSES

Before we can design the correction filter, we need to obtain the impulse responses, $\mathbf{h}_n$, for each observation point $\mathbf{r}_n$. To do this we play a suitable identification signal to the system and measure the acoustic waveform using a microphone at $\mathbf{r}_n$, as discussed in Sec. II. We must, however, acknowledge that the measurements are never error free; we will always have quantization errors, thermal noise, acoustic disturbances, etc. Also, we are using a discrete model of finite length, which never can describe the system perfectly. We have, therefore, chosen a Bayesian approach to estimate the impulse responses, which takes these uncertainties into account. More precisely, we have used the optimal linear estimator—also known as the linear minimum mean squared error (LMMSE) estimator or the maximum *a posteriori* estimator under Gaussian assumptions.

A. Prior information

A powerful feature in Bayesian analysis is that prior information is naturally incorporated, which significantly can improve our estimates. In this case we are interested in estimating acoustic impulse responses, and we know, for example, *a priori* that it takes a certain amount of time for the sound to travel from the active area of the loudspeaker to the observation point. The impulse response must consequently have a zero amplitude for that time interval, and data for this time interval can, therefore, be removed since it carries no information relevant to the design of the correction filter. For the response corresponding to the remaining time interval, which consists of a direct response and room responses (reflections and room modes), the direct response is often the strongest. However, if we have no knowledge from, for example, previous measurements, then it is difficult to say how much stronger the direct response is compared to the responses due to the room. Furthermore, we normally do not know if the impulse response amplitudes are more likely to be positive or negative. We can, however, say something regarding the mean signal variation since, in a typical measurement situation, the output power and the microphone gain is adjusted so that a sufficient acquired signal variation is obtained. This mean signal variation will correspond to a mean variation in the impulse responses, which we denote as

σ_h. Based on probability theory,[16,17] we can then assign a zero-mean Gaussian (prior) probability density function (pdf) to the impulse responses

$$p(\mathbf{h}_n|I) = \frac{1}{(2\pi)^{L_h/2}|\mathbf{C}_h|^{1/2}} e^{-(1/2)\mathbf{h}_n^T \mathbf{C}_h^{-1} \mathbf{h}_n}, \tag{5}$$

where $\mathbf{C}_h = \sigma_h^2 \mathbf{I} \, \forall \, n$.

B. The optimal linear estimator for acoustic impulse responses

To estimate the impulse responses $\mathbf{h}_n$, we first perform N experiments, one for each observation point, where we obtain the data $\mathbf{y}_n$, $n=1,2,\ldots,N$. Using the linear model (4), for the nth measurement point, the optimal linear estimate of $\mathbf{h}_n$ can be found by minimizing the mean squared error (MSE) criteria,

$$J(\mathbf{K}) = \arg\min_{\mathbf{K}} E\{\|\mathbf{h}_n - \hat{\mathbf{h}}_n\|^2\} = \arg\min_{\mathbf{K}} E\{\|\mathbf{h}_n - \mathbf{K}\mathbf{y}_n\|^2\}, \tag{6}$$

which has the closed form solution given by[18]

$$\hat{\mathbf{h}}_n = \mathbf{K}\mathbf{y}_n = \mathbf{C}_h \mathbf{U}^T (\mathbf{U}\mathbf{C}_h\mathbf{U}^T + \mathbf{C}_e)^{-1} \mathbf{y}_n, \tag{7a}$$

where $E\{\cdot\}$ is the expectation operator, $\mathbf{C}_e$ is the noise covariance matrix, $\mathbf{C}_h$ is the impulse response covariance matrix given by Eq. (5), and T is the transpose operator.

The covariance matrix $\mathbf{C}_h$ allows us to incorporate *a priori* knowledge regarding the impulse responses. A conservative setting is, as discussed in Sec. III A, to let $\mathbf{C}_h = \sigma_h^2 \mathbf{I}$; that is, we have no *a priori* knowledge of correlations and time varying amplitude variations of the individual magnitudes of the impulse response coefficients—if such correlations and amplitude variations do exist, then they must solely be given by the data.

The noise term, $\mathbf{e}$, in the model (4) above accounts for all uncertainties that we may have. As previously discussed we typically have model errors due to the finite length of the impulse responses, as well as the electronic measurement and the acoustic interference that cannot be avoided. Here we have also used a conservative model of the noise and let the covariance matrix have the form $\mathbf{C}_e = \sigma_e^2 \mathbf{I}$. The value of σ_e is, however, in practice seldom known beforehand. By using probability theory we could assign a prior for σ_e and integrate (marginalize) it out.[16,17] This would, however, lead to a more complicated non-Gaussian posterior. An often good approximation is to estimate σ_e from data and then use the estimated value in Eq. (7a), which is the method chosen in this paper (the procedure for estimating the noise variance is described in Appendix A). The estimator (7a) can be rewritten in the equivalent form

$$\hat{\mathbf{h}}_n = (\mathbf{U}^T \mathbf{C}_e^{-1} \mathbf{U} + \mathbf{C}_h^{-1})^{-1} \mathbf{U}^T \mathbf{C}_e^{-1} \mathbf{y}_n, \tag{7b}$$

which can be practically useful since the size of the matrix inverse may then be smaller. The latter form (7b) can also be obtained by assuming a Gaussian distributed $\mathbf{h}_n$ and $\mathbf{e}$ and then choosing the estimate $\hat{\mathbf{h}}_n$ where the *a posteriori* pdf has its maximum.

In summary, we have assumed that $\mathbf{h}$ and $\mathbf{e}$ are zero mean and mutually independent with covariance matrices $\mathbf{C}_h = E\{\mathbf{h}_n \mathbf{h}_n^T\} = \sigma_h^2 \mathbf{I}$ and $\mathbf{C}_e = E\{\mathbf{e}\mathbf{e}^T\} = \sigma_e^2 \mathbf{I}$, respectively.

C. The impulse response error

An important feature of Bayesian analysis is that we obtain a measure of the accuracy of our estimates. This is given by the posterior distribution, which for the estimator in Eq. (7a) and (7b) is Gaussian. The estimation error, $\boldsymbol{\epsilon} = \hat{\mathbf{h}} - \mathbf{h}$, is then also (zero-mean) Gaussian where the covariance matrix for $\boldsymbol{\epsilon}$ is given by[18]

$$\begin{aligned} \mathbf{C}_\epsilon &= \mathbf{C}_h - \mathbf{C}_h \mathbf{U}^T (\mathbf{U}\mathbf{C}_h\mathbf{U}^T + \mathbf{C}_e)^{-1} \mathbf{U}\mathbf{C}_h \\ &= (\mathbf{C}_h^{-1} + \mathbf{U}^T \mathbf{C}_e^{-1} \mathbf{U})^{-1}, \end{aligned} \tag{8}$$

where we have dropped $\mathbf{C}_e$'s dependence on $\mathbf{r}_n$ to simplify the notation.

IV. THE CORRECTION FILTER

The objective with the design of the correction filter is that the filter, as mentioned in Sec. I, should be robust in the sense that the listener should be able to move within a predefined region while still obtaining a reasonable correction. Since we are using a single source we cannot compensate for all (position-dependent) loudspeaker and room effects. That is, in particular, the reflected responses vary significantly over the region-of-interest,[5] and such effects, therefore, cannot be compensated—at least not without using a multiple-channel system. However, at lower frequencies the sound pressure varies more slowly with respect to listener movement, and if the position change is within a half wavelength then the sound wave will have a similar phase for the corresponding frequency. Thus, the correction filter can therefore compensate for the lower frequency room effects, and the cut-off frequency, where compensation is possible, is directly related to the size of our chosen correction region.[19] This information can be included in the design criteria, which will be further discussed below.

A. The linear minimum mean squared error correction filter

The objective is now to reproduce the input signal, $\mathbf{u}$, using linear pre-filtering according to

$$\hat{\mathbf{u}}_n * \boldsymbol{\delta}_{k-d_n} = \mathbf{h}_n * \mathbf{u} * \mathbf{f} + \mathbf{e} = \mathbf{H}_n \mathbf{U}\mathbf{f} + \mathbf{e}, \tag{9}$$

where $\mathbf{f}$ denotes the correction filter and where $\boldsymbol{\delta}_{k-d_n} = 1$ for $k = d_n$ and otherwise zero. Note that $\hat{\mathbf{u}}_n$ will be delayed relative to $\mathbf{u}$ due to the filtering delays associated with $\mathbf{h}_n$ and the filter delay due to the compensation filter $\mathbf{f}$ (i.e., it takes some time for the signal to pass through the system and the correction filter). We must, therefore, introduce a delay, d_n, in Eq. (9) in order to obtain a realistic estimator.

To find the linear filter we choose the filter, $\mathbf{f}$, that minimizes the mean squared sound reproduction error at a set of N observation points according to

$$\mathbf{f}_{\mathrm{OPT}} = \sum_{n=1}^{N} \arg\min_{\mathbf{f}} E\{\|\mathbf{u} * \boldsymbol{\delta}_{k-d_n} - \hat{\mathbf{u}}_n\|^2\}$$

$$= \sum_{n=1}^{N} \arg\min_{\mathbf{f}} E\{\|\boldsymbol{\Lambda}_{k-d_n}\mathbf{u} - \hat{\mathbf{u}}_n\|^2\}, \quad (10)$$

where $\boldsymbol{\Lambda}_{k-d_n}$ is the Toeplitz convolution matrix that corresponds to $\boldsymbol{\delta}_{k-d_n}$. We have again used a conservative approach and assumed that $E\{\mathbf{u}\mathbf{u}^T\} = \mathbf{C}_u = \sigma_u^2\mathbf{I}$.

Now, by using the estimated impulse responses in Eqs. (9) and (10) and equating the derivative with respect to $\mathbf{f}$ to zero, the correction filter which minimizes Eq. (10) becomes

$$\mathbf{f}_{\mathrm{LMMSE}} = \left(\sum_n \hat{\mathbf{H}}_n\hat{\mathbf{H}}_n^T\right)^{-1}\sum_n \hat{\mathbf{H}}_n^T\boldsymbol{\delta}_{k-d_n}, \quad (11)$$

where the dependence of the prior for $\mathbf{u}$ disappears since σ_u cancels out.

B. The LMMSE correction filter for room impulse responses with errors

Since we must estimate the room impulse responses from data, there will be deviations in our estimates from the true impulse responses, and, as discussed in Sec. III C, these deviations can be treated as Gaussian. A model that takes these uncertainties into account can then be expressed as

$$\mathbf{y}_n = (\hat{\mathbf{h}}_n + \Delta\mathbf{h}_n) * \mathbf{u} + \mathbf{e} = (\hat{\mathbf{H}}_n + \Delta\mathbf{H}_n)\mathbf{u} + \mathbf{e}, \quad (12)$$

where $\Delta\mathbf{h}_n \sim N(\mathbf{0}, \mathbf{C}_\epsilon^{(n)})$. Following the procedure used in Sec. IV A and using Eq. (12) in Eq. (10) results in

$$\mathbf{f}_{\mathrm{LMMSE}} = \left(\sum_n \hat{\mathbf{H}}_n^T\hat{\mathbf{H}}_n + \tilde{\mathbf{C}}_\epsilon^{(n)}\right)^{-1}\sum_n \hat{\mathbf{H}}_n^T\boldsymbol{\delta}_{k-d_n}, \quad (13)$$

where $\tilde{\mathbf{C}}_\epsilon^{(n)} = E\{\Delta\mathbf{H}_n^T\Delta\mathbf{H}_n\}$, which approximately is given by $\tilde{\mathbf{C}}_\epsilon^{(n)} \approx \text{Ł}_h\tilde{\sigma}_\epsilon^2\mathbf{I}$ (see Appendix B).

C. Reflection response filtering

Recall from the discussion in Sec. IV that we cannot compensate for reflections and high frequency room modes using a single source. Note, however, that since the reflection responses always arrive after the direct response, we can apply a suitable time varying filter to the impulse responses to attenuate the reflection effects in the impulse responses while keeping the low frequency room modes. Thus, by this procedure we can reduce the influence of the reflections in the design of the reconstruction filter.

The non-stationary filtering can be obtained by using a filter matrix where the columns are not delayed versions of the same filter, as for a normal Toeplitz convolution matrix, but now the filters are changing with time. This can be expressed in matrix form by multiplying the impulse response matrices $\mathbf{H}_n$ with a reflection filter matrix $\mathbf{F}_R$, that is, we replace $\mathbf{H}_n$ in the criteria (10) with $\mathbf{F}_R\mathbf{H}_n$, which results in the new criteria,

$$\mathbf{f}_{\mathrm{OPT}} = \arg\min_{\mathbf{f}} \sum_n E\{\|\mathbf{u} * \boldsymbol{\delta}_{k-d_n} - \mathbf{U}\mathbf{F}_R\mathbf{H}_n\mathbf{f}\|^2\}. \quad (14)$$

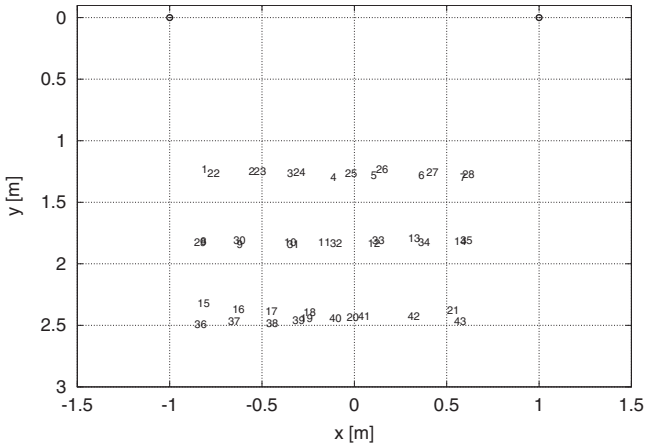


FIG. 2. The measurement setup. The circles indicate the loudspeaker positions and the numbers, 1–43, indicate the microphone measurement positions.

V. EXPERIMENTS

To verify the presented theoretical results above, experiments have been performed in a 5.8 m long, 4.5 m wide, and 2.6 m high room, which is designed after the live-end/dead-end principle; the wall behind the loudspeakers has sound absorbing draperies, and the wall at the opposite side to the loudspeakers has a sound diffuser. The data have been acquired using a DPA 4004 Reference microphone and an RME Multiface 24-bit sound card. The acoustic data were acquired at 43 measurement points, where measurements 1–21 were acquired at a height of 1.09 m and measurements 22–43 were acquired at 1.32 m, respectively. Two Genelec 8040A loudspeakers were used, and the layout of the experiments is shown in Fig. 2. In the results for the left loudspeaker presented here, the data from the 10th observation point, at $(-0.38, 1.83, 1.09)$ m, was used for evaluation, and the correction filter is based on data from the remaining 42 measurements. Furthermore, an impulse response (and correction filter) vector length of 16 384 elements was used here, which corresponds to a time duration of 371 ms. We have used white input signals of a duration of 3 s, sampled at a sampling frequency of 44.1 kHz, in the experiments, and the acoustic impulse responses were estimated using Eq. (7a) and (7b). We have given all estimated impulse responses an equal weight in the construction of the correction filter, and the data have, therefore, been normalized. We have also used a lower limit of 20 Hz and an upper limit 20 kHz for the working range of the correction filter.

In Fig. 3(a) the first part of the estimated impulse response for measurement point $(-0.38, 1.83, 1.09)$ m is shown. The response seen just before 2 ms is the direct path response, and the first (floor) reflection is the response seen at approximately 5 ms. One can also see that the frequency response, shown in Fig. 3(b), has many dips, which is known to cause problems in direct inversion approaches since then a division by zero nearly occurs.

Note that the arrival times of the same reflections but at different measurement points will not be the same since the propagation distances to the objects that reflect the sound will be different at different points. This is shown in Fig. 4(a)

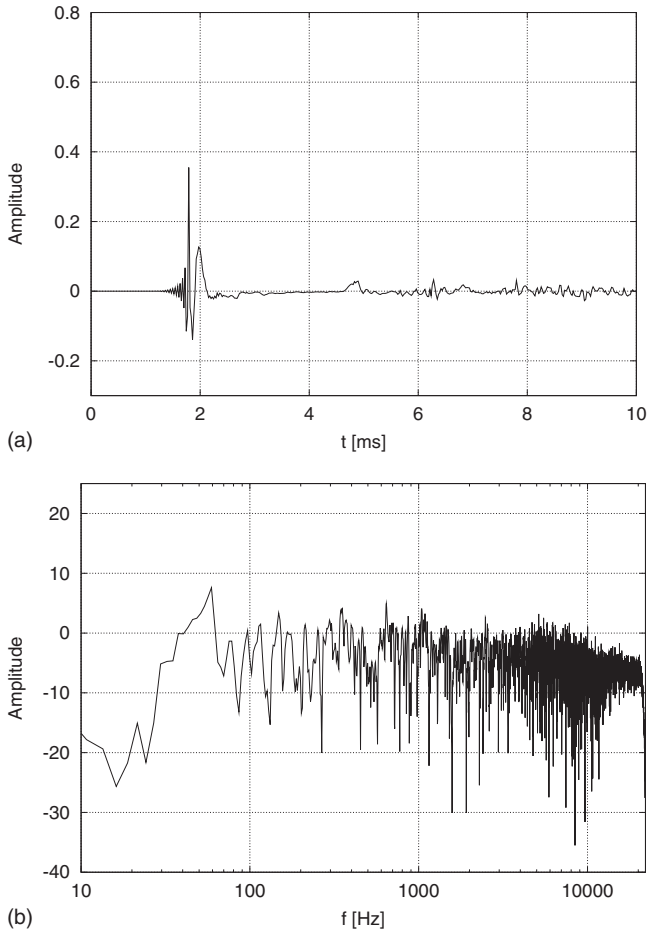


FIG. 3. The estimated acoustic impulse response for position (−0.38, 1.83, 1.09) (m).

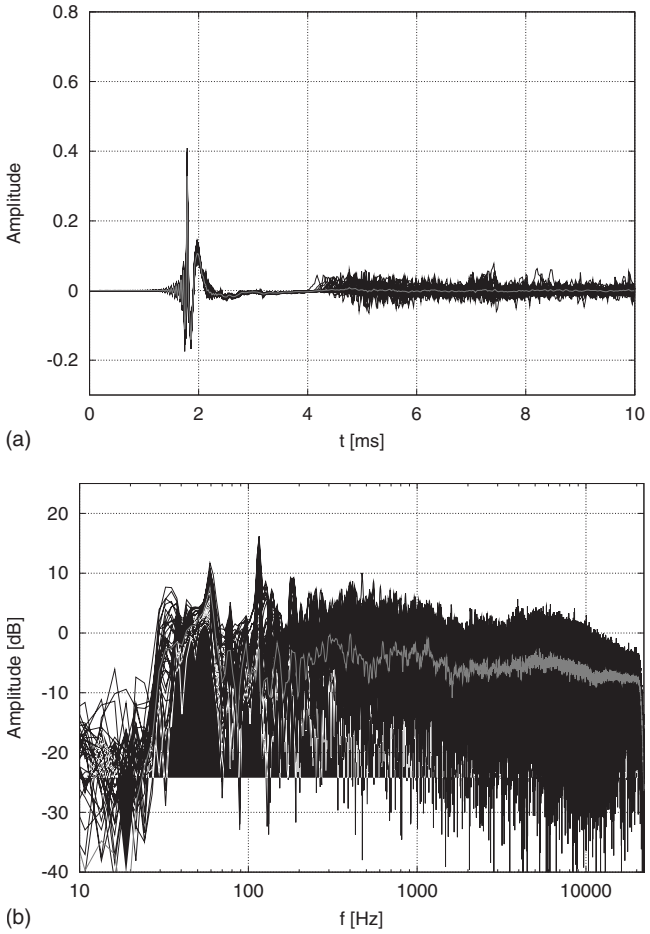


FIG. 4. The estimated acoustic impulse responses.

where all estimated impulse responses have been plotted in the same graph. As can be seen in the figure, the direct path impulse responses have similar, but not identical, shapes for all measurement points, but the reflections, which all arrive after 4 ms, vary significantly between different measurement points. This can also be seen by observing the average of the estimated impulse responses, shown as the gray solid line in Fig. 4(b); here it can be observed that the reflection part of the average impulse is close to zero since negative and positive responses does not arrive coherently for different points, and these parts will, therefore, almost cancel out in the average.

To further study the variability between estimated impulse responses, the spectra at each point have been plotted in the same graph in Fig. 4(b). Here we can see that the spectra have peaks and dips at roughly the same frequencies up to approximately 100–200 Hz, but above 200 Hz the peaks and dips occur at different frequencies for different points. This can also be seen in the average frequency response, shown as the gray solid line, which is relatively flat (with a falling trend) above 200 Hz with no strong isolated peaks or dips.

The LMMSE correction filter, based on the 42 "training" measurements, was then computed using Eq. (13), and the resulting filter is shown in Fig. 5. As seen in the frequency response in Fig. 5(b), the correction filter does not have excessive filter gain at any frequency. There is a peak at approximately 25 Hz since the loudspeakers do not transmit much energy below 30 Hz (see Fig. 4). One can also observe that there is three, or four, relatively well separated common resonance peaks in the spectra shown in Fig. 4(b) between 50 and 200 Hz. These peaks have corresponding dips in the correction filter's spectrum. One can also note that the correction filter compensates for the falling trend above 2–3 kHz seen in the average impulse response spectra in Fig. 4(b). The falling trend of the frequency responses can partly be explained by the diffraction effects of the acoustic source, which depend on the size of the source and the position of the measurement point; the fall-off trend for high frequencies increases when the distance from center-axis of the source increases (see, for example, Ref. 20).

To further show the properties of the LMMSE correction filter, the estimated impulse response for measurement point (−0.38, 1.83, 1.09) was filtered using the correction filter (shown in Fig. 5), which is displayed in Fig. 6. First one can observe, in Fig. 6(a), that the corrected system has a significantly shorter direct response compared to the uncorrected system. Second, the pre-ringings are roughly 50 dB lower than the maximum direct response [see Fig. 6(b)]. Third, the frequency response of the corrected system, shown in Fig. 6(c), is significantly closer to a flat response compared to uncorrected system. The correction is, as expected, not perfect since there are, in particular, still some dips in the spec-

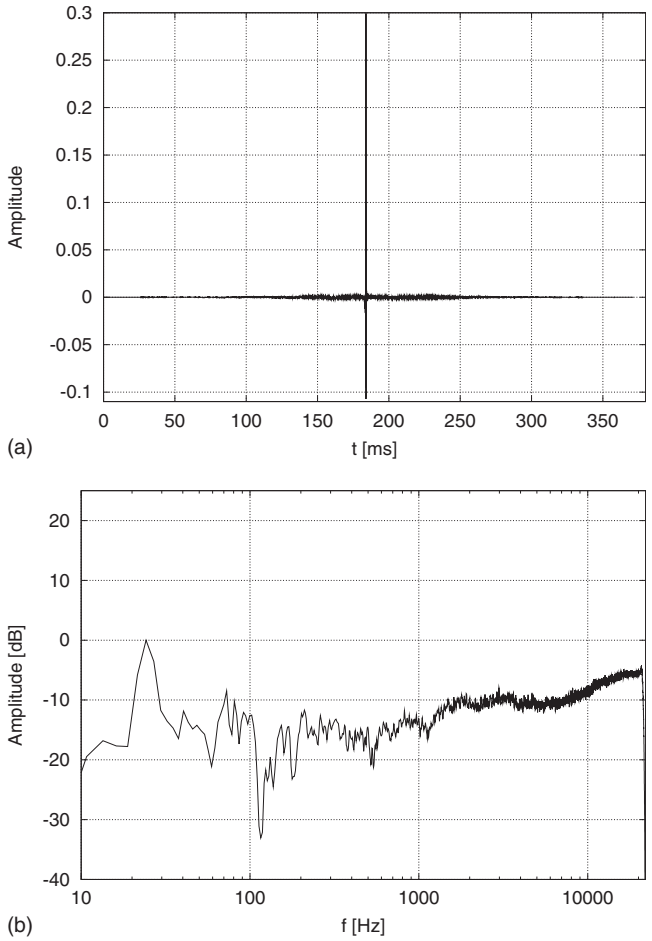


FIG. 5. The LMMSE correction filter.

tra for the corrected system. The high frequency part of the correction filter is also somewhat "noisy" [see Fig. 5(b)].

To see if the robustness of the correction can be improved by incorporating more prior information, we have also performed experiments using the (slightly heuristic) reflection filter approach described in Sec. IV C. Figure 7 shows the results where a non-stationary reflection filter with a cut-off frequency of 90 Hz was used when designing the correction filter. The start time for the non-stationary reflection filter was 3.6 ms, and the filter length was gradually increased until 5.6 ms where the full length (of four periods at the cut-off frequency) was reached. By comparing the results in Figs. 6(a) and 7(a), one can see that the correction of the direct response is similar both with and without reflection filtering, which was expected since the direct response was not affected by the filter. However, if we compare the results presented in logarithmic scale in Figs. 6(b) and 7(b), we can see that the reflection filter approach significantly has reduced the amplitudes of the pre-ringings. Furthermore, the high frequency parts of the correction filter is significantly smoother when reflection filtering is used, which can be seen by comparing Figs. 5(b) and 7(c).

VI. SUMMARY AND CONCLUSIONS

This paper treats the particular problem of the design of robust correction filters based on several impulse response

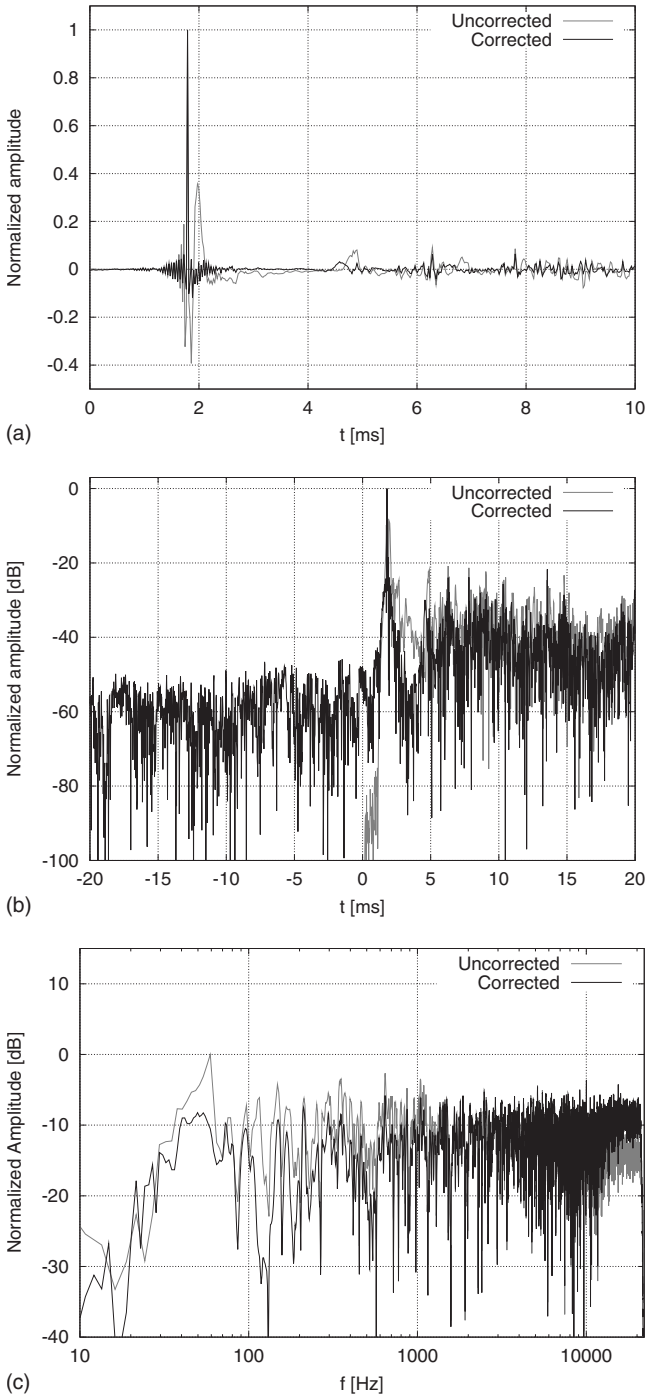


FIG. 6. The uncorrected and corrected impulse responses for measurement position $(-0.38, 1.83, 1.09)$ (m).

measurements. Clearly, the objective is to obtain the best correction possible, but we must first acknowledge the limitations of the experimental conditions and specify a realistic goal for our filter design. Here we are considering correction using a single source, and we are using a finite impulse response model where we must estimate the coefficients in the model from a limited number of noisy data. Due to the uncertainties, we have used a statistical approach in the design of the correction filter. The approach consists of two steps where we first estimate the impulse responses at a finite number of observation points in our region-of-interest, and then we construct a correction filter based on the estimated

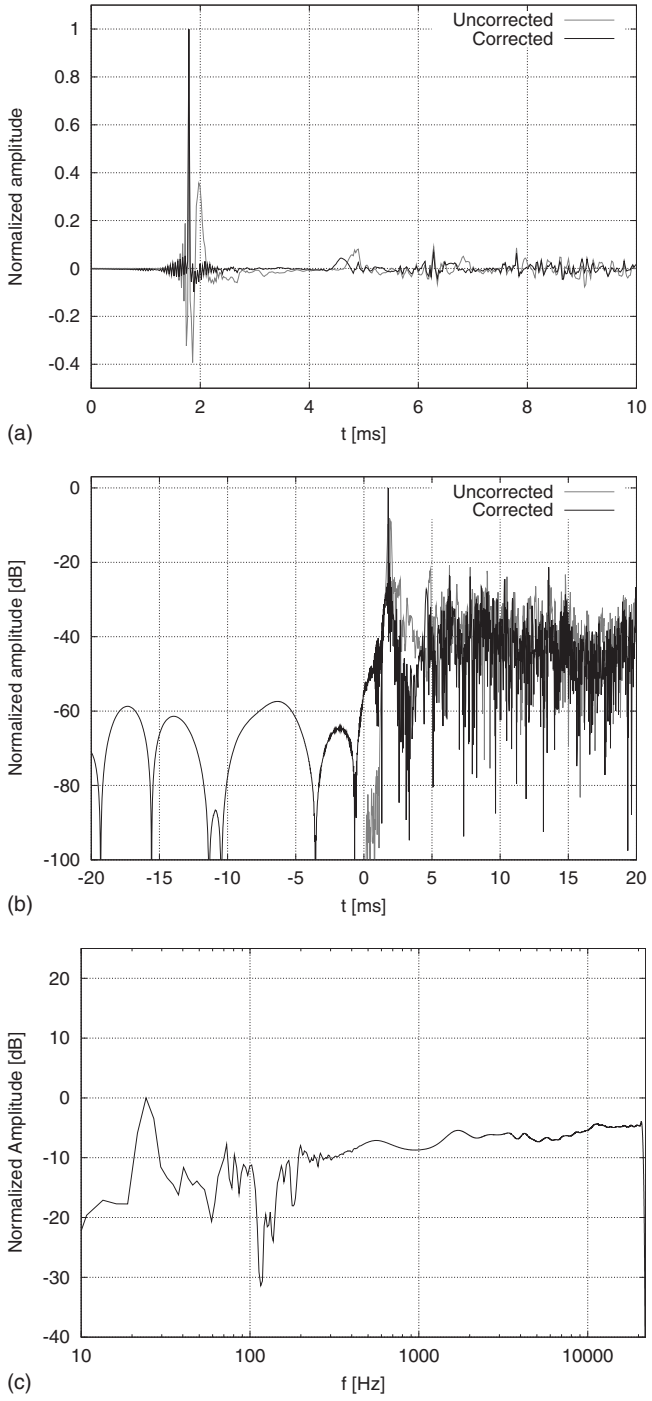


FIG. 7. The uncorrected and corrected impulse responses for measurement position (−0.38, 1.83, 1.09) (m) using a correction filter based on reflection filtered impulse responses.

impulse responses. By taking the uncertainties (i.e., the model error and measurement noise) into account, we have obtained reliable estimates of the impulse responses, which is vital for the design of the correction filter. The correction filter is then constructed by using a linear minimum mean squared criteria. We have, furthermore, introduced a time-dependent reflection filter in criteria, which attenuates the high frequency parts of the reflected responses, that is, the parts of the responses that we cannot compensate with a single source system.

As can be seen in the results from the experiments presented in this paper, the statistical multiple-point correction method introduced here is capable of significantly improving both the temporal and spectral responses compared to the uncorrected system. Furthermore, the obtained correction filter has a low level of pre-ringing, which was in the order of 60 dB lower than the direct response for the data used in this paper.

APPENDIX A: ESTIMATING THE NOISE VARIANCE

Here we are interested to estimate the noise variance σ_e^2 only and we, therefore, marginalize over the impulse response coefficients according to

$$p(\sigma_e|\mathbf{y}_n) \propto p(\sigma_e|I)\int_{\mathbf{h}_n} p(\mathbf{h}_n|I)p(\mathbf{y}_n|\mathbf{h}_n,\sigma_e,I)d\mathbf{h}_n = p(\sigma_e|I)\mathcal{L}(\sigma_e). \tag{A1}$$

If we, for convenience, use a uniform prior for $\mathbf{h}_n$, then the integral in Eq. (A1) becomes

$$\mathcal{L}(\sigma_e) = \int_{\mathbf{h}_n} \frac{1}{R_h}\frac{1}{(2\pi)^{L/2}\sqrt{|\sigma_e^2\mathbf{I}|}} \times \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y}_n-\mathbf{U}\mathbf{h}_n)^T(\mathbf{y}_n-\mathbf{U}\mathbf{h}_n)\right\}d\mathbf{h}_n, \tag{A2}$$

where $R_h=\mathbf{h}_{\max}-\mathbf{h}_{\min}$ is large enough so that the integral, in practice, can be evaluated from $-\infty$ to ∞. Then, after evaluating the Gaussian integral, Eq. (A2) becomes,

$$\mathcal{L}(\sigma_e) = \frac{1}{R_h}\frac{1}{(2\pi)^{L/2}\sqrt{|\sigma_e^2\mathbf{I}|}}\sqrt{\left(\frac{(\sigma_e^2 2\pi)^{L_h}}{|\mathbf{U}^T\mathbf{U}|}\right)} \times \exp\left\{-\frac{1}{2\sigma_e^2}\mathbf{y}_n^T(\mathbf{I}-\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T)\mathbf{y}_n\right\}. \tag{A3}$$

Furthermore, since the noise variance is a scale parameter, we use a (normalized) Jeffreys prior for σ_e,[17]

$$p(\sigma_e|I) = \frac{1}{\log(\sigma_{\max}/\sigma_{\min})}\frac{1}{\sigma_e}, \tag{A4}$$

where $\sigma_{\min}$ is a small value but larger than zero (typically determined by the bit resolution of the analog-to-digital converter) and $\sigma_{\max}$ is a large but finite value. Using Eqs. (A4) and (A3) in Eq. (A1), and equating the derivative of the log of the posterior (with respect to σ_e) to zero, then results in

$$\hat{\sigma}_e = \sqrt{\frac{\mathbf{y}_n^T(\mathbf{I}-\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T)\mathbf{y}_n}{L-L_h+1}}; \tag{A5}$$

that is, Eq. (A5) is the maximum *a posteriori* estimate of σ_e for a uniform prior of $\mathbf{h}_n$.

Recall from Sec. II that here the noise describes the model misfit, and different impulse response lengths will, therefore, result in different values of the estimated noise variance. As an example, the estimated σ_e for six impulse response lengths are given in Table I for the observation point at (−0.38, 1.83, 1.09). One can note that estimated σ_e becomes smaller when the impulse response length in-

TABLE I. Estimated σ_c for six different impulse response lengths, L_h, for the observation point (−0.38, 1.83, 1.09) (m).

L_h	512	1024	2048	4096	8192	16 384
σ_c	0.1644	0.1241	0.0585	0.0226	0.0126	0.0119

creases. This trend can also be seen in Fig. 8 where the normalized log likelihood for σ_e has been plotted for three different impulse response lengths. Here, one can also observe that the value of the most likely σ_e decreases when the impulse response length increases since a longer impulse response can describe the data better, and, furthermore, the posteriors are so narrow that they are essentially delta functions; hence, the actual limits of the prior for σ_e have negligible influence of the posterior.

APPENDIX B: AN APPROXIMATE EXPRESSION FOR THE ERROR COVARIANCE MATRIX

The error covariance matrix $\mathbf{C}_e$ was given in Eq. (8). This expression can be simplified by noting that since we are using a white excitation signal the matrix $\mathbf{U}^T\mathbf{U}$ will be nearly diagonal and can, therefore, be approximated with $\mathbf{u}^T\mathbf{u}\mathbf{I}$. By making this approximation, Eq. (8) reduces to

$$\mathbf{C}_\epsilon \approx \left(\frac{1}{\sigma_h^2}\mathbf{I} + \frac{\mathbf{u}^T\mathbf{u}}{\sigma_e^2}\mathbf{I}\right)^{-1} = \frac{\sigma_h^2\sigma_e^2}{\sigma_e^2 + \sigma_h^2\mathbf{u}^T\mathbf{u}}\mathbf{I}. \tag{B1}$$

If we use this approximation in the correction filter discussed in Sec. IV B, then the matrix $\tilde{\mathbf{C}}_e^{(n)} = E\{\Delta\mathbf{H}_n^T\Delta\mathbf{H}_n\}$ reduces to

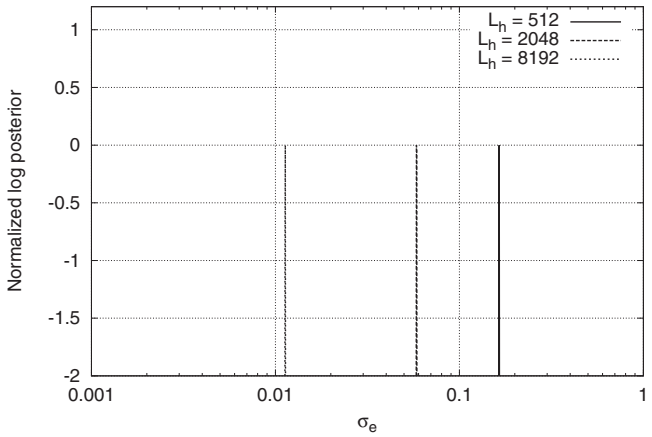


FIG. 8. The normalized log posterior for σ_e for three different impulse response lengths at observation point (−0.38, 1.83, 1.09) (m). The solid line corresponds to an impulse response length of 512 (right curve), the dashed line to 2048 (middle curve), and the dotted line to 8192 (left curve), respectively.

$$\tilde{\mathbf{C}}_\epsilon^{(n)} \approx \mathrm{L}_h\frac{\sigma_h^2\sigma_e^2}{\sigma_e^2 + \sigma_h^2\mathbf{u}^T\mathbf{u}}\mathbf{I} = \mathrm{L}_h\tilde{\sigma}_\epsilon^2\mathbf{I}, \tag{B2}$$

where we have dropped the dependence of n for σ_e^2 to simplify the notation.

[1] J. N. Mourjopoulos, "Digital equalization of room acoustics," J. Audio Eng. Soc. **42**, 884–900 (1994).
[2] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," J. Audio Eng. Soc. **47**, 583–595 (1999).
[3] L. D. Fielder, "Analysis of traditional and reverberation-reducing methods of room equalization," J. Audio Eng. Soc. **51**, 3–26 (2003).
[4] S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," J. Audio Eng. Soc. **37**, 899–907 (1989).
[5] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," IEEE Trans. Speech Audio Process. **8**, 311–319 (2000).
[6] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," J. Audio Eng. Soc. **48**, 259–280 (2000).
[7] Y. Haneda, S. Makino, and Y. Kaneda, "Multiple-point equalization of room transfer functions by using common acoustical poles," IEEE Trans. Speech Audio Process. **5**, 325–333 (1997).
[8] F. Asano, Y. Suzuki, and T. Sone, "Sound equalization using derivative constraints," Acust. Acta Acust. **82**, 311–320 (1996).
[9] S. Bharitkar and C. Kyriakakis, "A cluster centroid method for room response equalization at multiple locations," in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2001), pp. 55–58.
[10] R. Wilson, "Equalization of loudspeaker drive units considering both on- and off-axis responses," J. Audio Eng. Soc. **39**, 127–139 (1991).
[11] N. Stefanakis, J. Sarris, G. Cambourakis, and F. Jacobsen, "Power-output regularization in global sound equalization," J. Acoust. Soc. Am. **123**, 33–36 (2007).
[12] O. Kirkeby, P. A. Nelson, and H. Hamada, "Local sound field reproduction using digital signal processing," J. Acoust. Soc. Am. **100**, 1584–1593 (1996).
[13] L.-J. Brännmark and A. Ahlén, "Robust loudspeaker equalization based on position-independent excess phase modeling," in IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings, Las Vegas, NV (March 30–April 4, 2008).
[14] E. T. Jaynes, "Prior information and ambiguity in inverse problems," Inverse Probl. **14**, 151–166 (1984).
[15] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, UK, 1985).
[16] E. T. Jaynes, *Probability Theory: The Logic of Science*, 1st ed. (Cambridge University Press, Cambridge, UK, 2003).
[17] P. C. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge University Press, Cambridge, UK, 2005).
[18] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1993), Vol. 1.
[19] A. O. Santillán, "Spatially extended sound equalization in rectangular rooms," J. Acoust. Soc. Am. **110**, 1989–1997 (2001).
[20] F. Lingvall and T. Olofsson, "Statistically motivated design of input signals for modern ultrasonic array systems," J. Acoust. Soc. Am. **123**, 2620–2630 (2008).

Estimating the operating point of the cochlear transducer using low-frequency biased distortion products

Daniel J. Brown,[a] Jared J. Hartsock, Ruth M. Gill, Hillary E. Fitzgerald, and Alec N. Salt
Department of Otolaryngology, School of Medicine, Washington University in St. Louis, St. Louis, Missouri 63110

(Received 7 November 2008; revised 26 January 2009; accepted 26 January 2009)

Distortion products in the cochlear microphonic (CM) and in the ear canal in the form of distortion product otoacoustic emissions (DPOAEs) are generated by nonlinear transduction in the cochlea and are related to the resting position of the organ of Corti (OC). A 4.8 Hz acoustic bias tone was used to displace the OC, while the relative amplitude and phase of distortion products evoked by a single tone [most often 500 Hz, 90 dB SPL (sound pressure level)] or two simultaneously presented tones (most often 4 kHz and 4.8 kHz, 80 dB SPL) were monitored. Electrical responses recorded from the round window, scala tympani and scala media of the basal turn, and acoustic emissions in the ear canal were simultaneously measured and compared during the bias. Bias-induced changes in the distortion products were similar to those predicted from computer models of a saturating transducer with a first-order Boltzmann distribution. Our results suggest that biased DPOAEs can be used to non-invasively estimate the OC displacement, producing a measurement equivalent to the transducer operating point obtained via Boltzmann analysis of the basal turn CM. Low-frequency biased DPOAEs might provide a diagnostic tool to objectively diagnose abnormal displacements of the OC, as might occur with endolymphatic hydrops.
 [DOI: 10.1121/1.3083228]

PACS number(s): 43.64.Jb, 43.40.Le, 43.40.Ga, 43.38.Ar [BLM] Pages: 2129–2145

I. INTRODUCTION

There is a general consensus in the literature that tight regulation of the organ of Corti (OC) position and the related operating point (OP) of the cochlear transducer is necessary for the maintenance of hearing sensitivity, and modulation of the OP has been directly linked to auditory threshold fluctuations (Hirsh and Ward, 1952; Davis, 1983; Patuzzi *et al.*, 1984a, 1984b; Kim, 1986; Kemp, 1986; Dallos, 1992). Furthermore, it has been suggested that the hearing dysfunction in patients with Meniere's disease may be related to endolymphatic hydrops, which may involve an abnormal displacement of the OC (Valk *et al.*, 2004; Xenellis *et al.*, 2004). Presently there is no accurate, non-invasive objective measure of the OC position or OP. The main aim of the present study was to develop methods to estimate the OC position non-invasively using low-frequency biased distortion product otoacoustic emissions (DPOAEs). We also aimed to compare this estimate to measurements of the OP obtained through Boltzmann analysis of the cochlear microphonic (CM).

Researchers have previously estimated the OP using low-frequency tones to bias the OC position while measuring even-order distortion products in both the CM and DPOAEs (Frank and Kössl, 1996; Kirk *et al.*, 1997; Kirk and Patuzzi, 1997; Bian *et al.*, 2002; Bian, 2004; Sirjani *et al.*, 2004; Bian, 2006). When the OP of a nonlinear transducer with a first-order Boltzmann distribution is near zero (the symmetric point on the transfer curve), sinusoidal inputs produce relatively symmetric distorted outputs, and the amplitude of the even-order distortions in the output signal are relatively small compared to situations when the OP is displaced from zero and the output is asymmetric. Conversely, the amplitudes of odd-order distortions are largest when the OP is near zero and decrease as the OP is displaced from zero (Weiss and Leong, 1985; Frank and Kössl, 1996).

Despite recent research into the use of DPOAEs for the estimation of the cochlear transducer OP, there are many potential sources of nonlinearity in the cochlea that may contribute to DPOAE generation. One possibility is that they are generated by distortion of the mechanical-to-electrical transduction in the cochlea via the mechanoelectrical transduction (MET) channels on the hair bundles (Corey and Hudspeth, 1979), which is then reverse-transduced into mechanical vibration (Kim, 1986; Kirk and Yates, 1998; Yates and Kirk, 1998)), producing distortion products in the ear canal (Lukashkin and Russell, 1997, 1998; Liberman *et al.*, 2004). Another possibility is that DPOAEs are predominantly generated by distortion in the electrical-to-mechanical process that underlies the active amplification of OC vibration (Santos-Sacchi, 1991) or nonlinear compliance of the MET channels (Howard and Hudspeth, 1988; Jaramillo *et al.*, 1993). Whatever the mechanism underlying distortion in the cochlea, if the sources of distortion products in both the ear canal and the electrical response in the cochlea were the same, both sets of distortion products would have similar characteristics and might be used to determine the properties of transduction in the cochlea.

Forward-transduction in the cochlea follows a Boltzmann distribution due to the distribution of MET channel conductance (Holton and Hudspeth, 1986; Kros, 1996). Re-

[a]Author to whom correspondence should be addressed. Electronic mail: danielbrown@med.usyd.edu.au

searchers have used the CM as an approximation of the local MET channel conductance and have fitted a first-order Boltzmann function to the relationship between the CM and the acoustic stimulus (Corey and Hudspeth, 1983; Crawford *et al.*, 1989; Patuzzi and Rajan, 1990, Patuzzi and Moleirinho, 1998; Sirjani *et al.*, 2004), assuming that the CM is dominated by local hair cells (Dallos *et al.*, 1972; Patuzzi *et al.*, 1989) and that the MET channels are the dominant nonlinearity in the forward-transduction (Holton and Hudspeth, 1986; Kros, 1996). The three main parameters used to fit the Boltzmann function

$$y = V_{\text{off}} + \left(- P_{\text{sat}} + \frac{2 \cdot P_{\text{sat}}}{1 + \exp(2 \cdot S \cdot P_{\text{sat}}(\text{input} + \text{OP}))} \right)$$

are the OP, which is related to the asymmetry of transduction, the saturating voltage (P_{sat}), which is related to the potential between the open and closed states of the MET channels, and a sensitivity parameter (S), which represents the slope of the curve at the mid-point of the Boltzmann curve (Sirjani *et al.*, 2004). The OP is of particular interest because it is most likely a reasonable estimate of the resting conductance of the MET channels, which can be affected by a number of mechanical changes in the cochlea such as (1) displacements of the OC toward either scala vestibuli (SV) or scala tympani (ST), possibly due to a pressure differential across the partition (Nieder and Nieder, 1968a, 1968b; Patuzzi and Sellick, 1983; Russell and Kossl, 1992; Cheatham and Dallos, 1994), (2) hair cell length changes, which are a property of the outer hair cells (OHCs) (Brownell *et al.*, 1985; Zenner, 1986; Evans and Dallos, 1993), and (3) changes in the MET channel compliance (Howard and Hudspeth, 1988; Jaramillo *et al.*, 1993). Normally, the cochlea is relatively insensitive to dc pressures in the ear canal due to the acoustic high-pass filtering properties of the middle ear and helicotrema (Dallos, 1970; Ruggero *et al.*, 1986), and therefore the OP is most likely determined by cochlear factors, as mentioned above.

Despite the cochlear's insensitivity to dc pressures in the ear canal, we expected that low-frequency tones could be used to displace the OC position and transducer OP, thereby altering the phase and amplitudes of the distortion products in the CM and DPOAE evoked by a probe stimulus. We also predicted that properties of the transducer, such as its OP, could be indirectly determined from the amplitude and phase of the distortion products measured during low-frequency biasing. Presumably, the instantaneous pressure of a bias tone that causes even-order distortion products to be reduced to a minimum would represent the point where the transducer OP was displaced back to the mid-point of the Boltzmann transducer curve, which is what we define as zero OP. In this case, the applied bias pressure would be equal, but oppositely directed to the unbiased, innate OP of the transducer. This technique allows us to non-invasively probe the OP and monitor it during experimental manipulations.

II. METHODS

A. Animal preparation

Cochlear potentials and otoacoustic acoustic emissions (OAEs) were recorded in acute experiments in 19 NIH strain pigmented guinea pigs of either sex. Animals were anesthetized with an initial dose of 100 mg/kg sodium thiobutabarbital (Inactin, Sigma, St. Louis) and maintained in a state of deep anesthesia by periodic supplements via a cannula in the external jugular vein. Pancuronium bromide (2 mg/ml) was given intravenously as a muscle relaxant to reduce myogenic artifacts and to eliminate middle ear muscle contractions. Animals were artificially ventilated through a tracheal cannula, with end-tidal CO_2 maintained near 38 mm Hg (5%). Heart rate and vascular pO_2 were monitored with a pulse-oximeter (Surgivet, Waukesha, WI). Rectal temperature was maintained at 38 °C with a dc-powered, thermistor-controlled heating blanket. Animals were mounted in a head-holder, and the auditory bulla was exposed by a ventral approach, allowing access to the cochlea. The external ear canal was transected to allow placement of a hollow ear bar, through which acoustic signals were delivered.

The experimental protocols for this study were approved by the Animal Studies Committee of Washington University (Approval Nos. 20040209 and 20070147).

B. Acoustic stimulus generation, calibration, and acoustic emission recording

Tucker-Davis (TD) system-3 hardware was used to generate acoustic stimuli under the control of a custom-written Visual Basic (Microsoft) program. Three channels of sinusoidal stimuli were generated by two TD-RP2 modules, with each channel passed through a TD-PA5 attenuator to control level before being amplified by a headphone amplifier (TD-HB7). Two channels of probe stimuli (f1 and f2) were delivered to the ear canal by a modified Etymotic ER10C system coupled to the hollow ear bar. The third sound channel was used to generate and deliver 4.8 Hz bias tones. Bias tones were generated by a Sennheiser HD 580 or HD 265 driver, sealed into an acrylic coupler and connected to the ear bar with a short length of polyethylene tubing. All sound stimuli were delivered, and acoustic emissions were recorded within this closed sound system. The microphone within the ER10C system was used both for recording acoustic emissions and for the calibration of stimuli. In each animal, calibration curves were determined for all three transducers by tracking the attenuation required to generate a stimulus level of 70 dB SPL (sound pressure level) in the external canal as frequency was varied in $\frac{1}{4}$ octave steps. Calibration tables were generated, taking into account sensitivity changes of the microphone as a function of frequency. These calibration data were used by the software to deliver all subsequent stimuli calibrated on a dB SPL basis. Stimulus levels presented in this study were measured close to the tympanic membrane so they exclude the normal gain of 5–10 dB (from 2 to 10 kHz) provided by the ear canal (Sinyor and Laszlo, 1973). Calibrating at the tympanic membrane therefore means that the cochlea was exposed to stimulation levels of 5–10 dB lower than those experienced when stimuli are calibrated and pre-

TABLE I. Distortion levels measured in the ear bar, with and without an animal present. All levels are dB SPL, with standard deviation (±) and number of animals used (n) shown.

Stimulus	500 Hz, 90 dB SPL	500 Hz, 90 dB SPL	4 and 4.8 kHz, 80 dB SPL	4 and 4.8 kHz, 80 dB SPL
DPOAE:	2f1	3f1	f2-f1	2f1-f2
No animal (dB SPL)	35	18	−22	−4
Noise floor (dB SPL)	−15	−18	−27	−11
Animal average (dB SPL)	44 (±10, n=15)	42 (±7, n=15)	26 (±5, n=7)	29 (±10, n=7)
Noise floor (dB SPL)	−7 (±8)	−14 (±6)	−3 (±8)	−13 (±8)

sented free field. In order to calibrate the 4.8 Hz bias tones with the Etymotic ER10C microphone, it was necessary to extend the microphone response curve to lower frequencies, which was performed by comparing coupler responses from the ER10C with those of a Brüel & Kjær 4135 $\frac{1}{4}$ in. microphone as frequency was varied.

1. Transient hypoxia

Transient hypoxia was induced by coupling the ventilator output to the input through a container of CO_2 absorber (Sodalime, W.A. Grace, S.A.), so that the animal was rebreathing the same air, thereby slowly reducing the O_2 level, while the CO_2 was maintained at a constant low level to minimize vascular changes. The container incorporated a narrow vent line to maintain constant atmospheric pressure in the system. Gas sampling for CO_2 analysis was disabled during the procedure. Hypoxia was allowed to develop until endocochlear potential (EP) fell below approximately 50 mV, after which room-air ventilation was restored.

C. Cochlear potential recording (non-biased)

Cochlear potentials were recorded from a Ag/AgCl ball electrode placed on the bone at the margin of the round window (RW) membrane or from electrolyte-filled glass pipettes inserted into ST or scala media (SM). Signals from the RW electrode were recorded differentially with respect to a platinum needle electrode at the vertex using a TD HB4 optically coupled amplifier with a gain of 1000× and high-pass filtered at 5 Hz. Glass pipettes with a 5 μm tip diameter were filled with 500 mM NaCl for ST recordings or with 500 mM KCl for SM recordings. Responses from the glass pipettes were recorded through dc-coupled electrometer amplifiers with a gain of 10, referenced to the animal's earth electrode, which consisted of a salt-bridge AgCl electrode placed on the exposed muscles of the neck. EP was also recorded from pipettes inserted into SM, with the voltage measured prior to inserting the pipette into SM defining the zero level.

Cochlear potentials and acoustic emissions were collected simultaneously using the input channels of the TD-RP2 modules, sampling at 24 or 48 kHz. The software permitted data from four input channels to be collected and averaged simultaneously. In addition, two channels of data were also sampled at 44.1 kHz and streamed to disk using a PC equipped with an internal sound card. Time-averaged waveforms from the TD system were subjected to a number of analyses, including the following.

(i) Fast Fourier transform (FFT) spectral analysis using National Instruments Component Works routines called by the Visual Basic software. For responses collected without low-frequency biasing, FFT spectra were typically obtained from non-windowed time waveforms collected as the average of ten phase-locked 8192-point, 168 ms duration epochs. Three FFT spectra, each derived from independent data collections, were then averaged to form a final spectrum from which measurements were made. Spectral phase data were not averaged but were measured from the last block of data collected. The frequencies of probe stimuli were adjusted to ensure that the collected waveform contained an integer number of complete cycles, allowing sharp spectra to be obtained without windowing the time waveforms. The level of distortion generated by the sound delivery system, in response to one or two probe tones, was measured with the speakers and microphone placed in the ear bar, identical to experimental testing, but with the ear bar sealed into a 0.3 ml cavity, partly filled with wire wool to diffuse standing waves. A summary of the distortion product levels with and without an animal present is presented below in Table I. It should be noted that DPOAE levels in the animal depend greatly on the position of the OP in the cochlea where they are generated, which can vary between normally hearing animals, so there is an inherent complexity of averaging DPOAEs across animals. Nonetheless, we have provided average distortion levels for the purposes of indicating that they were well above that generated by the sound delivery system and well above the noise floor measured 10 Hz on either side of the distortion product.

(ii) A Boltzmann analysis of the CM waveform was performed using a 21 ms segment of the averaged time waveform described above. The waveform was transferred online to an Excel spreadsheet and fitted with a waveform synthesized from a Boltzmann curve relating voltage output to applied pressure input (Kirk *et al.*, 1997; Sirjani *et al.*, 2004). The parameters derived from this analysis were the following: P_{sat}: the saturation voltage of the curve, slope: the maximum steepness of the transducer function at zero input pressure, and OP_{Boltz}: the OP, which represents the position on the transducer curve at the zero crossings of the probe stimulus. Because the Boltzmann equation used in the present study multiplied the OP by the

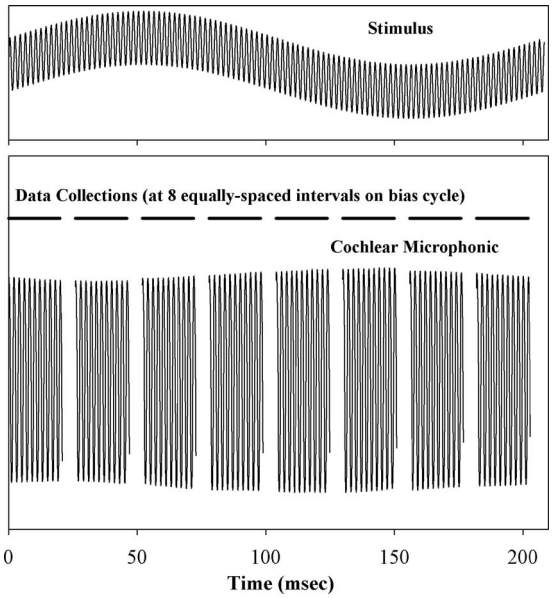


FIG. 1. Schematic of the experimental protocol in which a 4.8 Hz biasing tone is utilized to modulate OP of the cochlea during the simultaneous presentation of a probe tone. During each 208 ms cycle of the bias tone, eight independent collections were performed, and results were averaged over multiple bias cycles. Each of the eight data blocks was subsequently analyzed separately.

sensitivity to the probe stimulus [i.e., $2^{*}S/P_{\mathrm{sat}}{}^{*}$(input +OP)], the $\mathrm{OP}_{\mathrm{Boltz}}$ estimate obtained from CM measurements was effectively scaled in terms of probe tone pressure. Those parameters providing the best fit to a specific CM waveform were transferred back to the control program for plotting and storage. Online analysis of CM waveforms was only performed for responses to single-tone stimuli. Acoustic stimuli were turned on 5–6 s before data collection commenced to avoid onset phenomena, which might be related to efferent effects (Lukashkin and Russell, 2002; Halsey *et al.*, 2005).

(iii) The level of a tone-burst required to evoke a 10 μV compound action potential (CAP) waveform, averaged from ten consecutive responses evoked at a rate of 10.9/s, was measured for a range of tone-burst frequencies from 1 to 22 kHz. This allowed verification that the animal had normal sensitivity at the start of the experiment and that experimental protocols were not causing sensitivity changes.

D. Responses collected during low-frequency biasing

It has previously been shown that the guinea pig cochlea generates CM to stimuli with frequencies as low as 0.1 Hz (Salt and Demott, 1999), demonstrating that their OHCs are capable of transducing extremely low frequencies. The biasing protocol in the present study utilized a 4.8 Hz tone with a period of 208 ms, which allowed eight independent, equally-spaced, 21 ms data collections to be performed during each cycle of the bias (Fig. 1). The data buffers (1024 points with 48 kHz sampling or 512 points with 24 kHz sampling) were too short to obtain sharp spectra, so four independent data collections were made (32 buffers total), and the four waveforms collected at the same point on the bias were concatenated, producing 2048 points at 24 kHz or 4096 points at 48 kHz (84 ms of data in both cases), which was subjected to spectral and Boltzmann analysis. Stimulus frequencies were optimized to ensure an integer number of cycles in the original buffers so that concatenation did not contribute noise to the spectra. In addition, signals from the microphone were digitally high-pass filtered below 100 Hz to minimize transitions at the concatenation points. The eight independent time waveforms collected during the bias cycle were subjected to spectral and Boltzmann analysis as described above.

E. Modeled distortion generation

Interpretation of changes in DPOAE levels due to displacements of the OP was simplified by considering what was theoretically expected. The cochlear transducer was represented by a first-order Boltzmann transducer driven by one or two probe sinusoids, with OP varied to represent displacements by the bias tone. The generation of distortion products by a nonlinear transducer and their amplitude and phase relationship to the transducer's OP have been previously described using slightly different transducer properties as those used in the present study (Weiss and Leong in 1985; Frank and Kössl, 1996; Kirk *et al.*, 1997; Sirjani *et al.*, 2004). The specific parameters of the Boltzmann function used to model distortion generation were $y=1-2/[1+\exp(-4\cdot(x+\mathrm{OP}))]$, which were chosen so that the output waveform represented a typical CM recorded from the basal turn [Fig. 2(A), middle row]. The transfer function of the input sinusoid to the output waveform, with the OP at −1.2, zero, and +1.2, is shown in Fig. 2(B), where the input sinusoid plus the OP are plotted on the *x*-axis, and the output waveforms from Fig. 2(A) are plotted on the *y*-axis. The amplitude and phase of the distortion products measured from FFTs of the output waveform were then compared as OP was varied from one extreme (+1.2) to the other (−1.2) [Figs. 2(C)–2(J)].

Similar to the changes observed in previous studies, the amplitude of the 2f1 and f2-f1 distortion products (referenced to the amplitude of f1) was lowest when the OP was near zero [Figs. 2(C) and 2(G)], while the amplitude of the 3f1 and 2f1-f2 distortions was greatest when the OP was at zero [Figs. 2(E) and 2(I)]. The phase of the 2f1 and f2-f1 distortions flipped between 90° and 270° (relative to f1) as the OP changed from positive to negative values [Figs. 2(D) and 2(H)], while the phase of the 3f1 and 2f1-f2 distortion products was 0° or 180° depending on the OP value, flipping phase when the amplitude of these distortions was close to zero [Figs. 2(F) and 2(J)].

We also determined that when the transducer was driven by a probe sinusoid that produced substantial saturation of the transducer (i.e., when the amplitude of the output waveform was within 5% of maximum transducer output with the OP at zero), the 3f1 and 2f1-f2 distortions passed through a minimum as the OP was displaced. For one-tone probes, these minima occurred when the OP was near $\frac{1}{4}$ the peak-to-peak amplitude of the input sinusoid [Fig. 2(E)], while for two probe inputs these minima occurred near $\frac{1}{5}$ the peak-to-

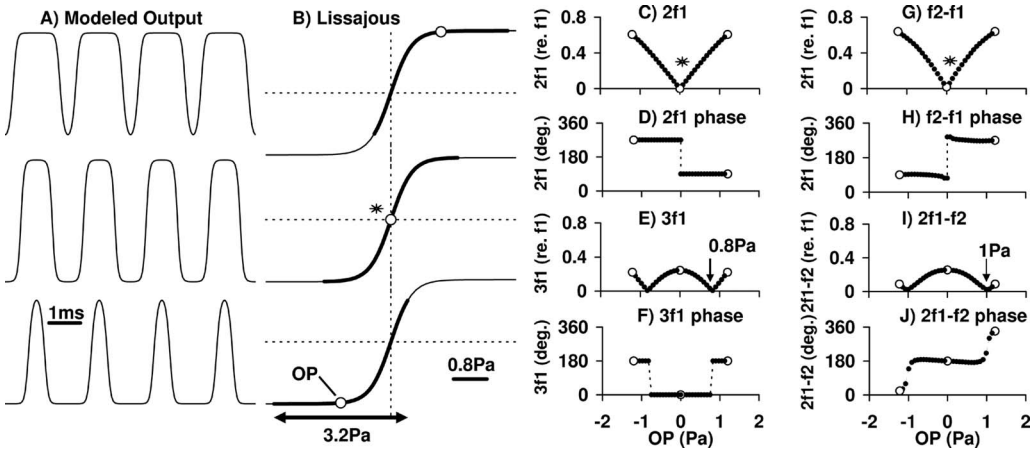


FIG. 2. Modeled changes in the distortion products generated by a pressure transducer with a first-order Boltzmann distribution driven by one or two sinusoids, as the transducer OP was varied. (A) The modeled transducer output when the OP was −1.2 Pa (top), 0 Pa (middle), and 1.2 Pa (bottom waveform), and the transducer was driven with a single 3.2 Pa peak-to-peak sinusoid. (B) The Lissajous curves of the modeled outputs from (A) (thick black line) plotted as a function of the sinusoidal input displaced by the OP (open circle) and approximated with a Boltzmann curve (thin black line). [(C)–(F)] The amplitude and phase of 2f1 and 3f1 distortion products (relative to f1) versus the OP, as the OP was biased from −1.2 to 1.2 Pa in 0.075 Pa steps. [(G)–(J)] The amplitude and phase of the f2-f1 and 2f1-f2 distortion products produced when the same transducer was driven with two sinusoids with a peak-to-peak amplitude of 5 Pa and the OP was varied from −1.2 to 1.2 Pa.

peak amplitude of probe sinusoids [Fig. 2(I)]. We found this relationship to be to be important because it allowed our OP_{Bias} measurement, derived from DPOAE biasing, to be referenced to the transducer displacement produced by the probe tone. This meant we could describe our OP_{Bias} measurement as a pressure relative to the probe tone (in pascal), by scaling our OP_{Bias} values so that these notches occurred at the appropriate pressures (presented in more detail in Sec. II F).

F. OP_{Bias} derived from the amplitude of the biased distortion products

With a clear understanding of the relationship between the OP and the amplitude distortion products, we could then begin to derive an estimate of the OP from the modulation of the DPOAEs throughout a bias tone. That is, in addition to estimating the Boltzmann OP (OP_{Boltz}) derived by fitting a synthesized probe stimulus to the CM waveforms, we could also derive an OP estimate from the bias-induced modulation of the distortion products measured from either the ear canal or CM, which we termed "OP_{Bias}." The OP_{Bias} estimate was derived by fitting a synthesized bias-induced modulation of distortion products generated by a first-order Boltzmann transducer (where the bias directly modulated OP of the transducer) to the measured distortion products in the ear canal or CM throughout the acoustic bias tone. While both the OP_{Boltz} and the OP_{Bias} are both estimates of the transducer OP and both are presented relative to the level of the probe tone(s), we have used different subscripts to clarify that the two measures come from completely different analytic techniques.

Figure 3 demonstrates how OP_{Bias} was derived from the eight measurements of f2-f1 and 2f1-f2 DPOAE amplitude throughout the 4.8 Hz bias. In cases where only one probe

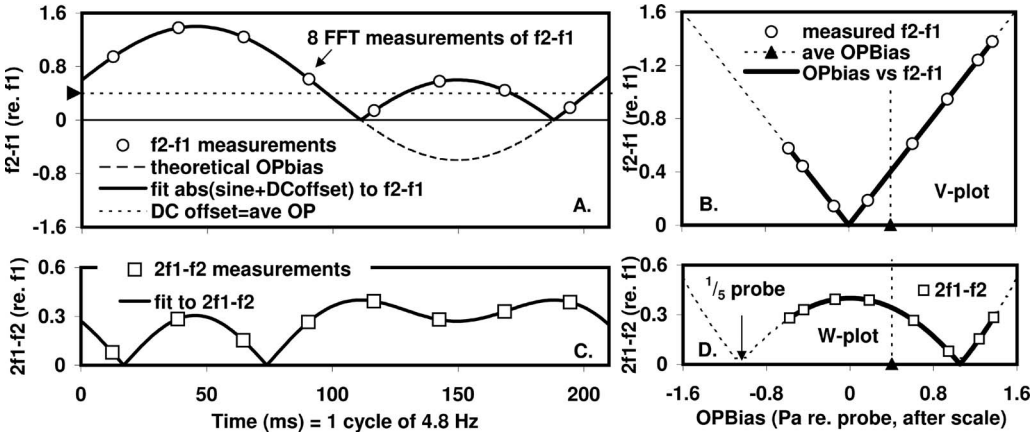


FIG. 3. (A) The eight FFT measures of f2-f1 throughout the bias (white circles), fitted with the absolute amplitude of a sinusoid plus a dc offset (thick black line). The dc offset was assumed to be proportional to the transducer OP without any bias (dashed line with arrowhead), and the sinusoid with the dc offset was assumed to be proportional to the modulation of the OP (dashed sinusoid). (B) The f2-f1 measurements and fit plotted against the sinusoid plus dc offset. (C) The eight FFT measurements of 2f1-f2 during the bias (white squares), overlaying the f2-f1 modulation. (D) The 2f1-f2 measurements plotted against the sinusoid plus dc offset fitted to the f2-f1 measurements, scaled so that the notches in the modulation (which was fitted with the absolute amplitude of an inverted parabola) occurred at $\frac{1}{5}$ of the peak-to-peak amplitude of the probe input (i.e., 5 Pa in this case).

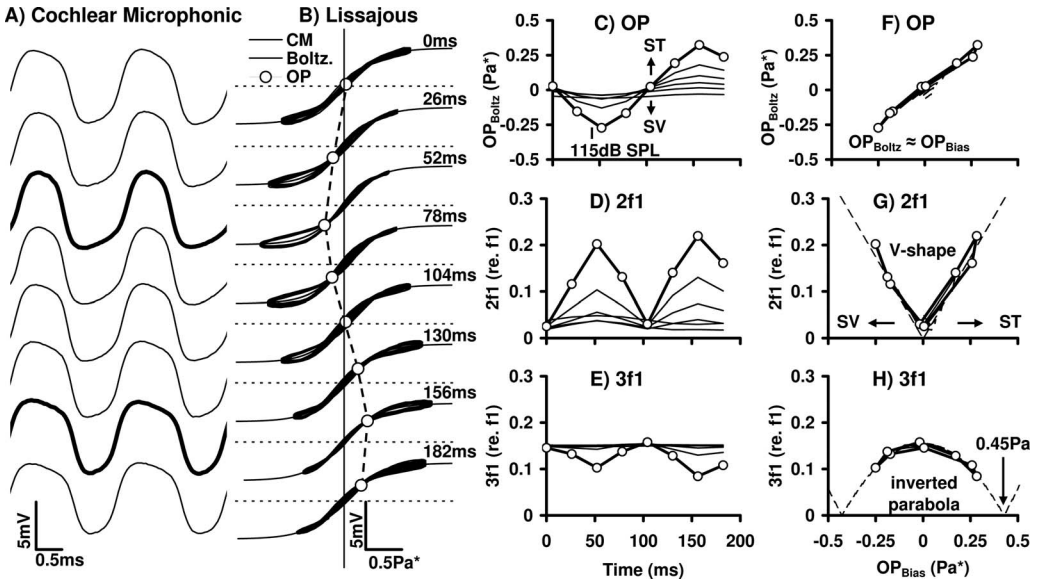


FIG. 4. Effects of a 4.8 Hz bias tone on the CM recorded from SM in the basal turn of animal PH05. (A) CM waveforms plotted for the eight points in the bias cycle as described in the methods. (B) CM data from (A) (black, coursing lines) versus the sinusoidal input displaced by the fitted OP_{Boltz} (white circles), shown with the fitted Boltzmann transducer curve (extended black line). [(C)–(E)] The OP_{Boltz} and the amplitudes of 2f1 and 3f1 respectively plotted as a function of time during one bias cycle. The amplitudes of the distortion products (2f1 and 3f1) are presented as a ratio relative to f1. [(F)–(H)] OP, 2f1, and 3f1 respectively plotted as a function of OP_{Bias}. The near-linear relationship between the OP_{Boltz} and OP_{Bias} values throughout the bias suggests that OP_{Bias} was as accurate an estimate of the transducer OP as the OP_{Boltz} measure. Furthermore, there was a V-shaped relationship between 2f1 and the OP, and a W-shaped relationship between 3f1 and the OP [dashed lines in (G) and (H)], with 3f1 nearing zero as the OP neared $\frac{1}{4}$ the pressure of the probe tone (equal to 0.45 Pa for the 500 Hz, 90 dB SPL tone).

tone was used, the OP_{Bias} was derived from the modulation of 2f1 and 3f1 using the same method. The same technique was also used to derive OP_{Bias} from the bias-induced modulation distortion products in CM waveforms. In the figure, the absolute amplitude of a 4.8 Hz sinusoid with an added dc offset was fitted to the spectral amplitude of eight f2-f1 distortion product measurements (relative to the spectral amplitude of f1), which were modulated by the low-frequency bias [Fig. 3(A)]. The dc offset, amplitude, and phase of the sinusoid were adjusted to best fit the modulated amplitude of the f2-f1 distortion product. The instantaneous "sinusoid +dc offset" at each point on the bias waveform was theoretically proportional to the transducer OP throughout the bias, and the amplitude of f2-f1 could be plotted against it to produce a *V*-plot [Fig. 3(B)]. In order to directly compare the OP_{Bias} and OP_{Boltz} estimates, we transformed OP_{Bias} onto a scale related to the pressure of the probe tone. To scale OP_{Bias}, the amplitude of the biased 2f1-f2 (relative to f1) was plotted against the bias sinusoid+dc offset pressures, as was done for f2-f1 [Fig. 3(D)]. This resulted in a "W-shaped" curve, similar to an inverted parabola, with minima in the amplitude of the distortion products that occurred at specific bias values. From our model of distortion generated by a first-order Boltzmann function (Sec. II E), we determined that these minima theoretically occurred at $\frac{1}{4}$ of the peak-to-peak amplitude of the probe for single-tone inputs or $\frac{1}{5}$ peak-to-peak amplitude for two-tone inputs [Figs. 2(E) and 2(I)]. We could therefore scale our OP_{Bias} values so that notches in the modulated 2f1-f2 distortion product occurred at values which were equivalent to $\frac{1}{5}$ the peak-to-peak pressure of the probe. If there was insufficient modulation to produce a minimum in 2f1-f2, we extrapolated the modulation using an inverted parabola fitted to the modulation pattern. The scaled dc offset [Fig. 3(A)] or the average of all OP_{Bias} points fitted throughout the bias cycle [Figs. 3(B) and 3(D), vertical dashed line] represented the transducer OP at the zero crossings of the bias, i.e., comparable to measurements without a bias present.

III. RESULTS

A. Low-frequency biasing of the cochlear microphonic

An example of the effect of a 4.8 Hz 115 dB SPL bias tone on the CM evoked by a 500 Hz 90 dB SPL tone, recorded from the SM in the basal turn, is shown in Fig. 4(A). The two heavy traces show the responses nearest the maximum and minimum bias pressures, in which considerable differences in CM waveshape are apparent. The thick coursing lines on the Lissajous plots [Fig. 4(B)] show the same CM waveforms (*Y*-axis) plotted relative to the 500 Hz sinusoidal input (*X*-axis) with the bias-induced OP changes added to it. The extended thin lines are the fitted Boltzmann functions, and the white circles indicate the OP_{Boltz}, which is the location on the Boltzmann curve at the zero crossings of the 500 Hz input sinusoid. The changes in the OP_{Boltz} produced by the bias tone almost entirely account for the bias-induced shape changes of the CM. The bias-induced changes in OP_{Boltz}, 2f1, and 3f1 for the eight time points during the bias are shown in Figs. 3(C), 3(E), and 3(D), respectively. Curves with symbols show changes during the 115 dB SPL bias tone, while thinner lines show the similar but progressively smaller modulations as the bias tone was reduced in 5 dB steps.

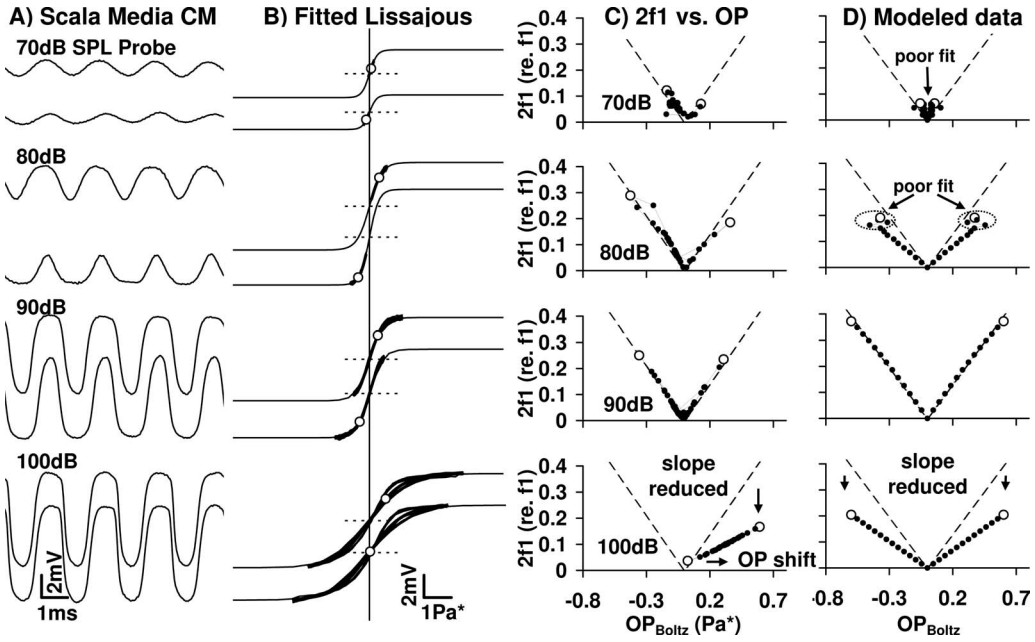


FIG. 5. (A) The CM waveforms from animal PH09, evoked by 70, 80, 90, and 100 dB SPL, 500 Hz tones, at the extremes of the 115 dB SPL, 4.8 Hz bias, with the level of the 500 Hz probe shown above each set of two CM waveforms. (B) The Lissajous curves of the CM waveforms from (A) (coursing black line) plotted as a function of the probe displaced by the fitted OP_{Boltz} (open circles) and approximated with a Boltzmann curve (extended black line). (C) The amplitude of the 2f1 distortion products (relative to f1) versus the OP_{Boltz} derived from the CM evoked by different level 500 Hz probes, recorded throughout 4.8 Hz bias tones at 100, 105, 110, and 115 dB SPL. The dashed lines in (C) is the V-shaped relationship between 2f1 and the OP_{Boltz} when the stimulus was 90 dB SPL. (D) The modeled 2f1 versus OP_{Boltz} plots when a modeled first-order Boltzmann transducer was driven with different level sinusoids, and the OP was varied from −0.6 to +0.6. The outputs of the modeled transducer were ac coupled before being re-fitted with a Boltzmann function to mimic the ac coupling of the CM responses. The extreme OP displacements are shown as white circles.

Changes in OP_{Boltz} approximately followed a sinusoidal modulation during the 4.8 Hz sinusoidal bias tone, demonstrating that OP_{Boltz} was a reliable indicator of the transducer input. The relative amplitude of the second harmonic (2f1, relative to the amplitude of f1) showed a bimodal modulation, with distortion minima occurring when OP_{Boltz} was near zero and maxima when OP_{Boltz} was furthest from zero. Conversely, the relative amplitude of the third harmonic (3f1, relative to f1) was minimal when OP_{Boltz} was biased from zero and maximal when OP_{Boltz} was at zero. The three measures plotted against OP_{Bias} are shown in Figs. 4(F)–4(H). OP_{Boltz} followed a near-linear relationship with OP_{Bias} [Fig. 4(F)], suggesting that both methods estimated the transducer OP similarly. The second harmonic showed substantial changes in amplitude as the OP was displaced (93% modulation), producing a V-shaped relationship when plotted against OP_{Bias} [Fig. 4(G)]. Changes in the 3f1 amplitude throughout the bias were far smaller (35% modulation), with the amplitude of 3f1 largest when OP_{Bias} was closest to zero and reaching a minimum as OP_{Bias} neared $\frac{1}{4}$ the peak-to-peak pressure of the probe (0.45 Pa).

B. Single-tone probe level and modulation of the OP

The probes used to evoke CM or DPOAEs in this study were high-level, sufficient to produce saturated CM waveforms without being traumatic. CAP thresholds and the EP, measured before and immediately after the presentation of our DPOAE probes, were unaltered by the presentation of probes at levels lower than 100 dB SPL for 500 Hz tones and lower than 85 dB SPL for 4 and 4.8 kHz frequencies (Salt *et al.*, 2009).

Examples of biased CM waveforms produced by 500 Hz probe tones at 70, 80, 90, and 100 dB SPL, at the extreme bias points during a 115 dB SPL 4.8 Hz tone, are shown in Fig. 5(A). At each level of probe stimulus, CM waveforms for the positive and negative limits of the bias are plotted against time [Fig. 5(A)] or against the input stimulus [Fig. 5(B)]. On the Lissajous curves in Fig. 5(B), the output waveforms are indicated by thick dark lines and the OP is indicated by open circles. At each probe level, the difference between the two paired curves demonstrates the OP changes induced by the bias.

The amplitude of 2f1 from the CM evoked at each probe level used was measured during a range of 4.8 Hz bias tones at levels of 100, 105, 110, and 115 dB SPL and is plotted versus the OP_{Boltz} in Fig. 5(C) [i.e., each plot in Fig. 5(C) consists of a group of 2f1 versus OP_{Boltz} measurements, with the probe level indicated on the left, but with all bias tone levels]. Calculated changes in 2f1 and OP_{Boltz} fitted to the ac-coupled output of a first-order Boltzmann equation (using the same equation described in Sec. II E), driven by different level probes and with various OP displacements, are also shown in Fig. 5(D). Initially, it appeared that there was less modulation of the OP by the bias tone when the 500 Hz probe level was 70 dB SPL. However, we found the same result in both the CM and model-derived waveforms and demonstrated that it was due to errors in fitting a Boltzmann curve to a relatively undistorted waveform evoked by a low-

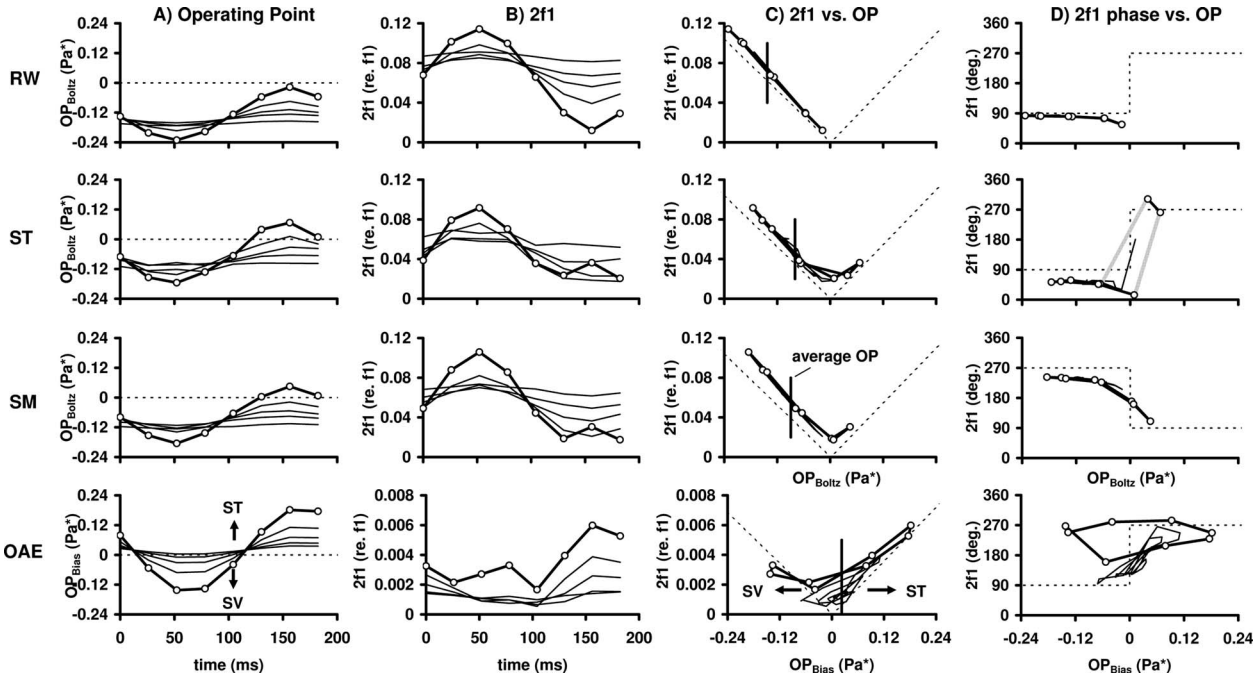


FIG. 6. [(A) and (B)] The OP_{Boltz} and 2f1 (white circles) simultaneously measured in animal PH10 from the RW CM (top row), ST CM of the basal turn (second row), SM CM of the basal turn (third row), and the OP_{Bias} and 2f1 from the ear canal OAE (bottom row) during simultaneous presentation of both 500 Hz, 90 dB SPL and 4.8 Hz, 115 dB SPL tone. The OP_{Boltz} from the CMs were derived from Boltzmann fits to the CM waveforms, while the OP_{Bias} was derived from the relative displacement of the DPOAEs throughout the 4.8 Hz bias tone, as described in Sec. II F. Thin black lines are similar modulations using lower sound-level bias tones from 95 to 110 dB SPL in 5 dB steps. [(C) and (D)] The amplitude and phase of 2f1 versus the OP_{Boltz} or OP_{Bias} for each recording location. The dashed lines in each plot are the modeled changes in the amplitude and phase of 2f1 for similar OP modulations.

level stimulus. To avoid inaccurate Boltzmann fits to the CM, we generally used probe levels that produced sufficiently saturated CM waveforms, without producing any temporary or permanent changes in cochlear function, which in the case of 500 Hz tones was around 90 dB SPL.

Apart from an underestimation of the OP with low-level tones, the slope of the V-shaped relationship between 2f1 and the OP_{Boltz} [Fig. 5(D), dashed lines] varied for different input levels, particularly for 100 dB SPL. We attributed this to a variation in the relative amplitude of 2f1 to f1 over the range of input levels that was also present in the modeled data. One main difference between the real [Fig. 5(C)] and modeled [Fig. 5(D)] changes in the relationship between 2f1 and the OP_{Boltz} was a substantial positive shift in the average OP_{Boltz} in the real data when the probe level was 100 dB SPL, representing a physiological adaptation of the OP toward ST during the high-level probe tone. This OP displacement with high-level probes has been reported in previous studies (Patuzzi and Moleirinho, 1998).

C. Simultaneous modulation of distortion products from different locations

In a number of experiments, we simultaneously estimated the OP from different recording locations during 4.8 Hz biasing. The estimated OP_{Boltz} derived from the CM recorded from the RW, ST of the basal turn, and SM of the basal turn and the estimated OP_{Bias} derived from the modulated DPOAE are shown throughout the low-frequency bias in Fig. 6(A). Symbols show the influence of the bias tone at 115 dB SPL with thin lines representing similar modulations at lower bias levels from 110 to 95 dB SPL in 5 dB steps. All OP measures were modulated in-phase and by a similar amount, except for OP_{Bias} from the DPOAEs, which was modulated 40% more than any of the OP_{Boltz} measures. The average OP throughout the bias differed between recording locations [Fig. 6(C), vertical lines], although the average OP from ST and SM measurements, which were at the same region along the length of the cochlea, was similar. This resulted in different modulations of the second harmonic, e.g., 2f1 from the RW was modulated sinusoidally, while 2f1 from ST, SM, and the ear canal measurements was modulated bimodally. This was best explained when the amplitude of 2f1 was plotted against the modulated OP from each recording location [Fig. 6(C)]. In these plots there was a V-shaped relationship indicated by the dashed line in each plot. As the OP from the ST, SM CM, and the OAE recordings passed through zero during the bias, the 2f1 was bimodally modulated. In contrast, the mean OP estimated from the RW CM was at a more negative value and the displacement by the bias did not cause it to cross zero, so the 2f1 exhibited a simpler sinusoidal modulation. The phase of 2f1 from each location is plotted against the OP for each set of data in Fig. 6(D). As predicted from the model (see Fig. 2), 2f1 phase changed by $180° \pm 40°$ when the OP crossed zero for the ST and SM recordings. However, the same phase relationship was not observed in the 2f1 versus OP_{Bias} plot from DPOAE [Fig. 6(D), bottom plot] because the phase of f1 in the ear canal was dominated by the stimulus and therefore was unrelated to the phase of the distortion products generated within the cochlea.

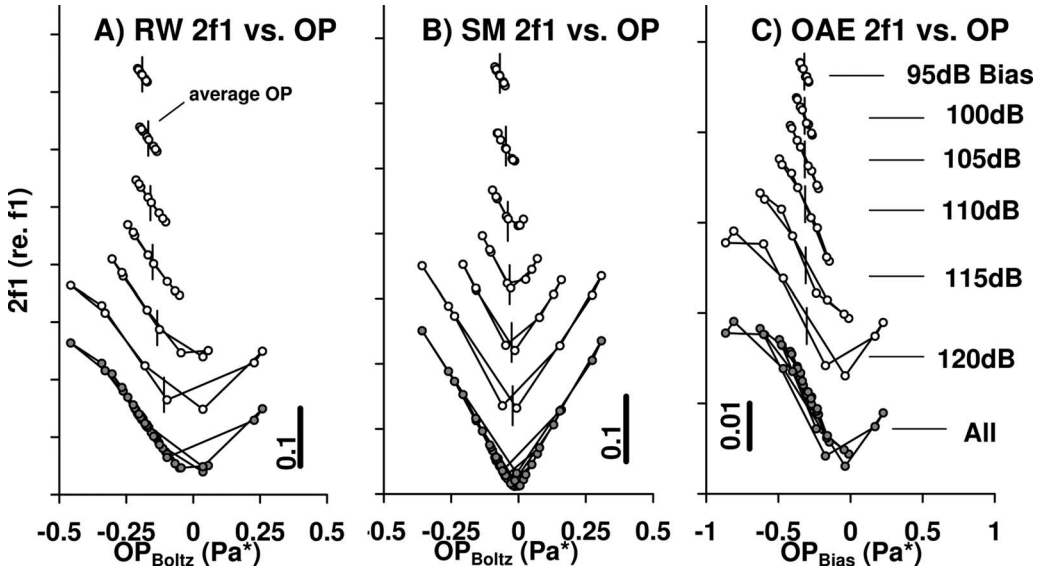


FIG. 7. Animal PH13. Modulations of the OP_{Boltz}, OP_{Bias}, and 2f1 with different levels of the 4.8 Hz tones ranging from 95 to 120 dB SPL (shown at the right of the figure). In each plot, 2f1 (relative to f1) has been plotted versus the estimated OP for each of the bias tone levels (open circles), with each bias level then shifted vertically from the next. In each plot, all biases within that plot have been overlaid at the bottom (gray circles). [(A) and (B)] 2f1 versus the OP_{Boltz} derived from the RW CM and SM CM evoked by a 500 Hz, 90 dB SPL tone during the 4.8 Hz bias. (C) 2f1 DPOAE evoked by a 500 Hz, 90 dB SPL tone versus OP_{Bias} derived from the modulation of the DPOAEs throughout the 4.8 Hz bias.

D. Bias level with a single-tone probe

We intended to use the average OP_{Bias} value as an estimate of the resting OC displacement and transducer OP. Therefore, we investigated if the bias tone altered the *average* OP by measuring the OP_{Boltz} and OP_{Bias} during the bias tone at a range of bias levels from 95 to 115 dB SPL. The average OP throughout the bias only underwent minor changes as the level of the bias tone increased [e.g., 0.08 Pa, Fig. 7(A), dashed vertical lines]. Higher-level bias tones typically produced larger modulations of the OP and 2f1, maintaining the V-shaped relationship, with minimal changes in the average OP values. In each plot of Fig. 7 the amplitude of 2f1 versus the OP at each bias level has been shifted vertically for clearer visualization of the overall OP bias (Fig. 7, white circles), with all bias levels then overlaid at the bottom of each plot (gray circles). The level of bias for each plot is provided at the right of the figure. The plots of 2f1 versus OP_{Boltz} derived from the RW and SM CM evoked by a 500 Hz, 90 dB SPL probe tone are shown in Figs. 7(A) and 7(B), while 2f1 versus the OP_{Bias} from the simultaneously recorded DPOAE is shown in Fig. 7(C). In all three cases the average OP was negative, equivalent to a displacement of the OC toward SV. At lower bias levels the 2f1 was modulated sinusoidally, while at higher bias levels (at 115 dB for the RW, 105 dB for SM, and 115 dB for the OAE), the 2f1 was modulated bimodally.

As in Fig. 7, higher-level biases produced proportionally more modulation of the OP up to 120 dB SPL, which was the highest bias level used in this study. The average modulation of the OP_{Boltz} derived from the CM evoked by a 500 Hz, 90 dB SPL tone, for different bias levels, recorded from both the RW and SM is shown in Fig. 8. For bias levels above 80 dB SPL, modulation of OP_{Boltz} (on a logarithmic scale) increased in a linear fashion with increasing bias level, demonstrating that the displacements produced by the bias tones were not saturating over this range for levels up to 120 dB SPL. Typically, levels of the 4.8 Hz bias lower than 80 dB SPL produced modulations of the OP that were below the normal variation in the OP_{Boltz} throughout the bias, i.e., representing the noise floor (Fig. 8, white circles).

E. Bias level with a two-tone probe

Figure 9 shows changes in f2-f1 distortion (upper row) and 2f1-f2 distortion (lower row) measured at different sites,

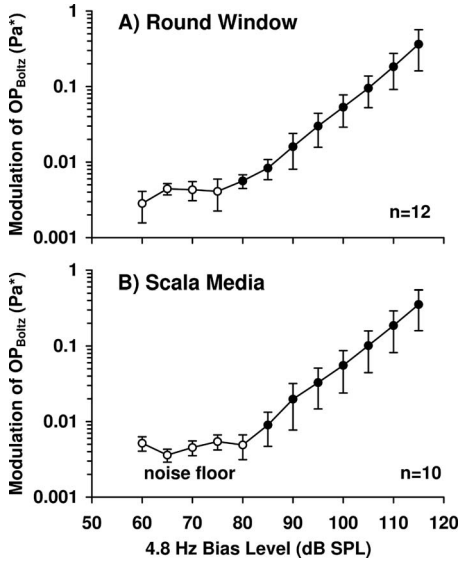


FIG. 8. The average modulation of the OP_{Boltz} (peak-to-peak OP change) derived from the 500 Hz, 90 dB SPL evoked CM recorded from (A) the RW and (B) SM of the basal cochlear turn, with the level of the 4.8 Hz bias varied from 60 to 115 dB SPL. The number of animals used to obtain the average modulation is shown in each figure. White circles represent modulations that were below the average variation in the OP_{Boltz} throughout the bias (the noise floor).

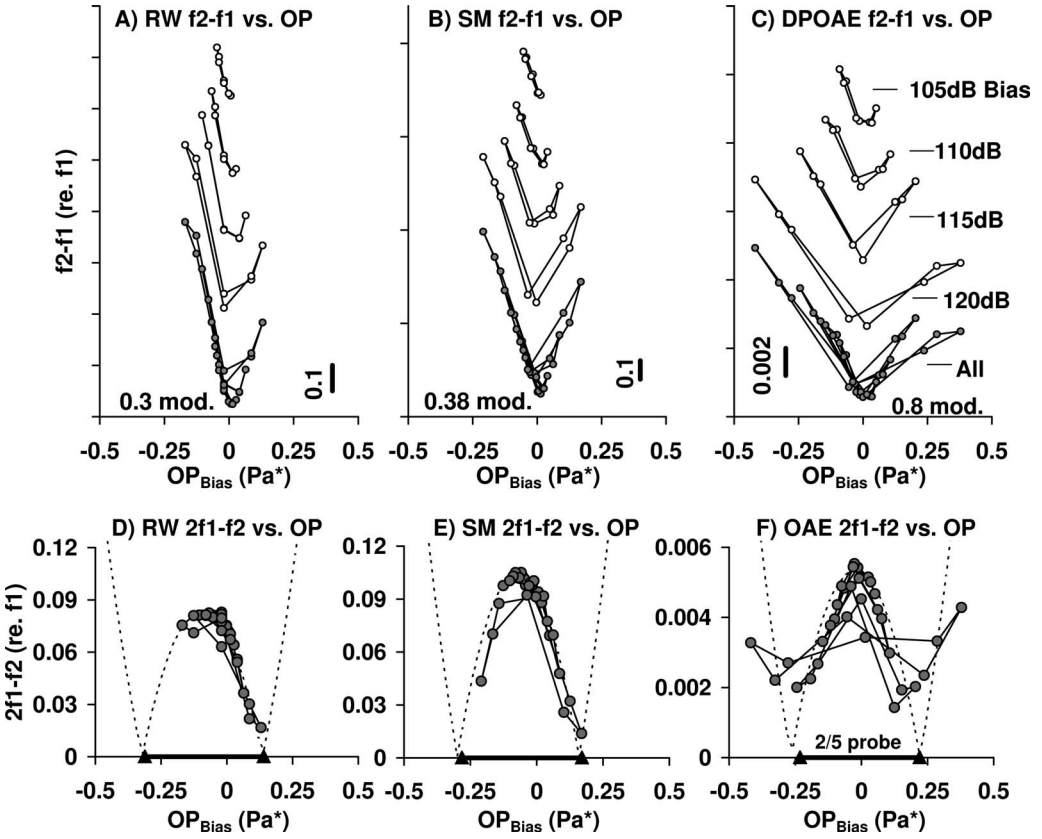


FIG. 9. Animal PH13. Modulations of the OP_{Bias} and f2-f1 derived from the modulated RW CM, SM CM, and DPOAE responses evoked by 4 and 4.8kHz, 80 dB SPL tones during different levels of a 4.8 Hz tone, ranging from 105 to 120 dB SPL (shown at the right of the figure). [(A)–(C)] f2-f1 (relative to f1) versus the OP_{Bias} for each of the bias tone levels (open circles), with each bias level shifted vertically from the next, and all biases within that plot overlaid at the bottom (gray circles). [(D)–(F)] 2f1-f2 versus the OP_{Bias} derived from each recording, demonstrating that the amplitude of 2f1-f2 neared zero when the OP_{Bias} was displaced to $\frac{1}{5}$ of the peak-to-peak pressure of the probe or that the difference between the OP_{Bias} values at which the notches occurred was $\frac{2}{5}$ the pressure of the probe.

as bias level was varied from 105 to 120 dB SPL. The probe stimuli in this case were tones at 4 and 4.8 kHz presented at 80 dB SPL. With the high-frequency probe stimuli, a reliable Boltzmann analysis of the CM was not possible [due to factors such as electrical low-pass filtering of the CM by the cochlear scalae; see Strelioff (1973) and Patuzzi *et al.* (1989)], and therefore OP_{Boltz} could not be obtained. In contrast, our estimate of OP_{Bias} is unaltered by changes in the overall amplitude of the distortion products (as might occur with electrical or acoustical filtering) but depends upon their modulation and the appearance of notches or minima throughout the bias. This allowed us to estimate OP_{Bias} even though Boltzmann analysis of the CM was not possible. We have therefore plotted the distortion products from each of the RW, SM, and OAE recordings against OP_{Bias} for each location. The f2-f1 distortion product from each recording was modulated bimodally [Figs. 9(A)–9(C)], particularly at higher levels of bias pressure in a similar manner to the modulation of 2f1 evoked by single tones (Fig. 7). The magnitude of OP_{Bias} derived from the ear canal was over twice that derived from the basal turn CM recordings evoked by the two-tone stimulus, suggesting that the bias produced more displacement of the OC at the region where the DPOAEs were generated in the cochlea than at the base of the cochlea where the local CM was recorded.

The amplitudes of the 2f1-f2 distortion products are shown in Figs. 9(D)–9(F), further illustrating the technique used to relate OP_{Bias} values to the probe. The OP_{Bias} values were scaled so that minima in the modulated 2f1-f2 occurred 0.45 Pa apart (which was 2/5 the peak-to-peak pressure of the probe stimulus in the ear canal). It was necessary to extrapolate the minima in the case of the RW and SM recordings [Figs. 10(D) and 10(E)], while the DPOAE 2f1-f2 was modulated sufficiently to produce notches [Fig. 10(F)]. While our modeled data suggested that the amplitude of 2f1-f2 should have been maximal when the amplitude of f2-f1 was close to zero, there was a slight difference between the OP_{Bias} values at the maximal 2f1-f2 position and the minimal f2-f1 position, particularly for the RW recording. These responses are analogous to the bias-induced modulation of 3f1 to a single tone [Fig. 4(H)] and confirm that the OP can be derived from ear canal DPOAE measurements.

F. Using distortion measures to track OP during hypoxia

One way of comparing OP_{Boltz} from CM and OP_{Bias} from DPOAEs is to compare the two measures during manipulations that are known to change OP. Figure 10 shows the results from an experiment in which OP_{Boltz} and OP_{Bias} measures from the CM and DPOAE were compared during hypoxic episodes. Hypoxia has been shown to produce

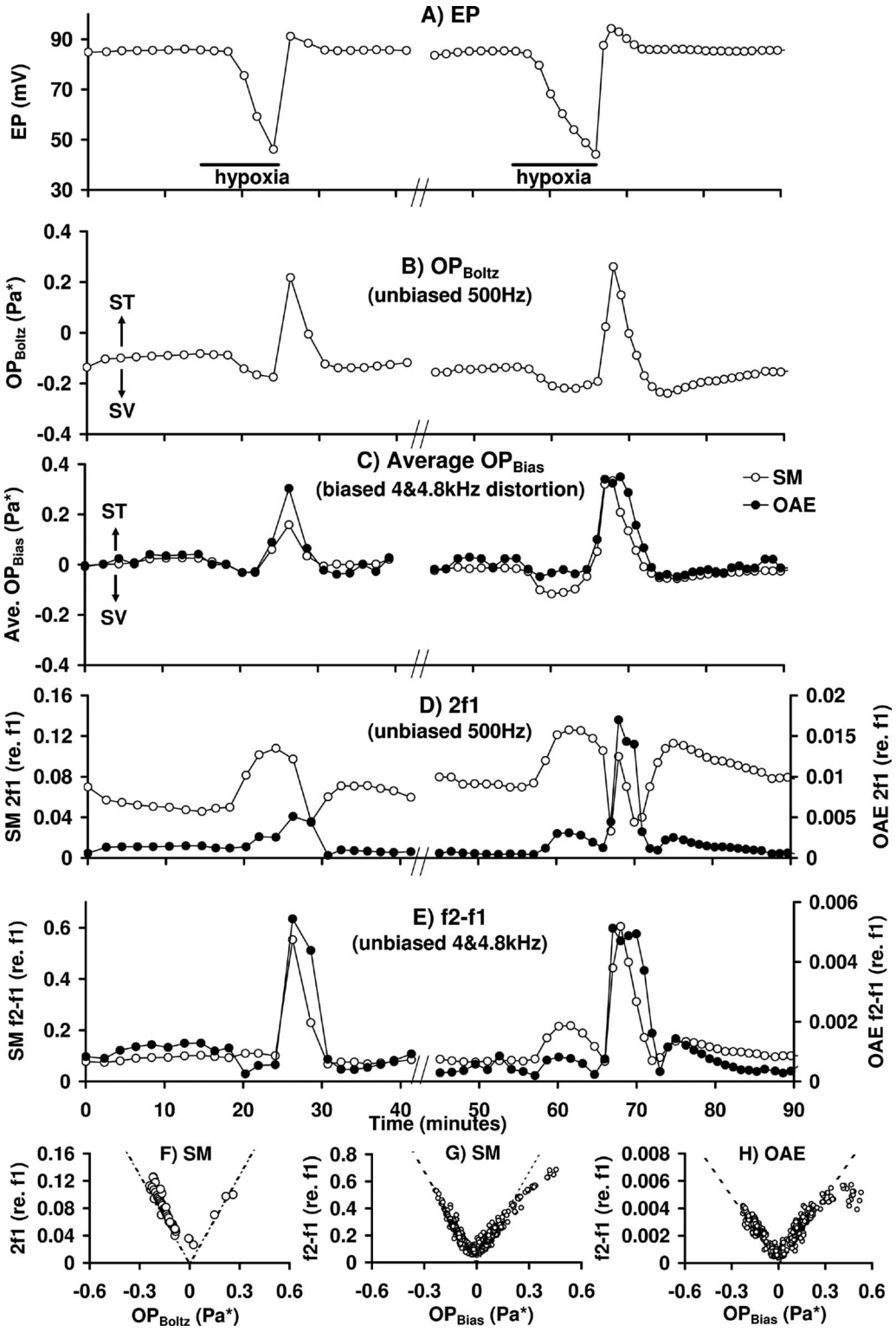


FIG. 10. [(A)–(E)] Repeated measurements in animal PH14 during two hypoxic events [rebreathing times indicated by dark bars in (A)], with a 20 min break between recordings (indicated as line-breaks). (A) EP. (B) OP_{Boltz} obtained from 500 Hz, 90 dB SPL evoked CM. (C) The average OP_{Bias} obtained from biased CM and DPOAE responses evoked by 4 and 4.8 kHz 80 dB SPL tones during a 4.8 Hz, 115 dB SPL tone. (D) The relative amplitude of 2f1 from the CM and DPOAE evoked by a 500 Hz, 90 dB SPL tone. (E) The relative amplitude of f2-f1 from the CM and DPOAE evoked by 4 and 4.8 kHz, 80 dB SPL tones with no bias. Black circles represent measures from the CM, and white circles those from the ear canal. (F) A *V*-plot of the data in (B) versus the data in (D). [(G) and (H)] *V*-plots of the data from (E) versus the data in (C). Note that there are eight times more points in (G) and (H) than in (F) because there were eight measures of the distortion products and OP_{Bias} during the low-frequency bias.

changes in the EP, DPOAEs, and Boltzmann parameters (Rebillard and Lavigne-Rebillard 1992; Patuzzi and Moleirinho, 1998; Olzowy *et al.*, 2008). Responses were repeatedly evoked by successive probes consisting of a 500 Hz, 90 dB SPL tone, simultaneously presented at 4 and 4.8 kHz, 80 dB SPL tones with a 115 dB SPL, 4.8 Hz bias tone and at 4 and

4.8 kHz, 80 dB SPL tones without a bias tone, with approximately 10 s of silence separating each probe set. Biased responses were used to derive the average OP_{Bias} from both the ear canal and the CM, while responses without the bias provided the unbiased two-tone distortion products.

Around 3–5 min after hypoxia was initiated, the EP began to decline gradually [Fig. 10(A)]. Concurrently, the OP_{Boltz} and OP_{Bias} values shifted toward SV [Figs. 10(B) and 10(C)]. When EP reached 48 mV normal ventilation was restored and the EP rapidly increased with an over-shoot before recovering to around 85 mV. As the EP recovered, OP_{Boltz} and OP_{Bias} shifted to around +0.2 Pa, equivalent to a shift toward ST, before decreasing with a slight over-shoot toward negative values. After a 20 min break in recordings, a second hypoxia was induced, producing similar but slightly larger changes in each measurement. During this second hypoxia CM, EP, and distortion products were monitored more frequently.

Prior to the hypoxic events, the OP_{Bias} was closer to zero than OP_{Boltz}. This difference was most likely due to differences in probe stimuli rather than being a difference in the method of estimating the OP *per se*. Regardless of this initial difference, the induced changes in OP_{Bias} and OP_{Boltz} were similar throughout both hypoxic events.

The amplitudes of the unbiased 2f1 and f2-f1 distortion products from the 500 Hz, 4 kHz, and 4.8 kHz probes throughout the recordings are presented in Figs. 10(D) and 10(E). While there were differences in the amplitudes of the unbiased 2f1 distortion product between CM and OAE recordings, this can be explained on the basis of a different initial OP of the transducer responses dominating the CM and the DPOAEs. The amplitudes of the unbiased f2-f1 distortion product from the CM and DPOAE were similar throughout the experiment, suggesting that the transducer generating f2-f1 in the CM and DPOAE had similar initial OPs. Induced distortion changes are more easily explained using plots of 2f1 or f2-f1 versus the OP_{Boltz} or OP_{Bias}, as shown in Figs. 10(F)–10(H). In all cases, distortion was reduced as the OP measure moved closer to zero.

IV. DISCUSSION

The present study outlined the dependence of DPOAE amplitude on the OP of the transducer and demonstrated that with a clear understanding of this relationship, DPOAE measurements can be used to non-invasively monitor cochlear function. In particular, low-frequency biasing of DPOAEs could be used to non-invasively estimate the resting position of the OC, which might provide a clinical tool for diagnosing disorders that perturb the normal transducer position, such as endolymphatic hydrops. Previous studies have attempted to use the amplitude of the DPOAEs as indicators of cochlear function (Rebillard and Lavigne-Rebillard 1992; Kossowski *et al.*, 2001; Kujawa and Liberman, 2001; Garner *et al.*, 2008; Olzowy *et al.*, 2008), reporting complex changes in the DPOAEs that were difficult to interpret, particularly as changes in the OP were not taken into consideration. Furthermore, other studies have taken similar approaches to investigate hearing function in humans (Mrowinski *et al.*, 1996; Hensel *et al.*, 2007; Bian and Scherrer, 2007) and animals (Frank and Kössl, 1996; Kirk *et al.*, 1997; Lukashkin and Russell, 2002; Bian, 2004; Sirjani *et al.*, 2004; Salt *et al.*, 2005) and have reported similar relationships between pressure in the ear canal, the OP, and the amplitude of the DPOAEs as those reported here. However, this study is the first to directly relate the OP obtained from CM recordings to the OP derived solely from simultaneously recorded DPOAEs.

A. Differences between the OP_{Boltz} derived from CM and the OP_{Bias} derived from DPOAEs

While both OP_{Boltz} derived from the CM and OP_{Bias} derived from the DPOAEs were modulated sinusoidally during the low-frequency bias and both changed in a similar manner during the hypoxia experiments in the present study, often there were differences between the average OP_{Boltz} and average OP_{Bias}. As discussed by Kirk *et al.* (1997), who directly compared the Boltzmann OP from basal CM recordings to the biased DPOAEs, there are several possible reasons why there might be differences between the OP_{Boltz} and the minimal point of the f2-f1 DPOAE throughout the bias (equivalent to the average OP_{Bias} value in our study). To briefly summarize the points made by Kirk *et al.* (1997); first, the nonlinear transducer generating the DPOAEs might be better described by a *second-order* Boltzmann distribution, resulting in differences between the OP_{Boltz} (which was estimated using a *first-order* Boltzmann function) and our measurement of the OP_{Bias}. Second, there might be differences in adaptation to the probe tone between sources underlying the generation of OP_{Boltz} and OP_{Bias}. We present another explanation here, that the OP_{Boltz} derived from the basal turn CM is an estimate of the OC position at the base of the cochlea, while the OP_{Bias} derived from the DPOAEs is an estimate of the OC position at the overlapping region stimulated by the probe tone(s), more apical in the cochlea, and that there were differences between the resting OC position at either location. Therefore, it would be unreasonable to directly relate the amplitude of the DPOAEs to the OP_{Boltz} or distortion products from basal turn CM recordings, and it is essential to apply an individual OP estimate to the transducer generating the DPOAE.

The portion of the cochlea the OP_{Boltz} and OP_{Bias} measurements represent is most likely a weighted sum of the regions stimulated by the probe tones contributing to the CM and DPOAE measurements and most likely increases with the level of the stimulus as broader regions of the cochlea are recruited. That the OP_{Boltz} derived from the RW CM was often different from the OP_{Boltz} derived from ST or SM recordings in the present study (Fig. 6), which were as little as 4 mm apical from the RW, further suggests that the OC position can differ throughout the cochlea and that the OP_{Boltz} estimate is a relatively local measure of OC position. Furthermore, that the estimates of OP derived from ST and SM measurements made at the same region along the length of the cochlea were similar, even though their CM amplitudes differed almost tenfold, and their polarities were opposite,

indicates that the OP differences between locations throughout the cochlea were not due to the analysis technique.

Previous studies of DPOAEs suggest they are generated at both the "overlapping" region of the cochlea stimulated by the two probe tones (near the characteristic region of the probes) *and* at the characteristic region of the DPOAE itself (Talmadge *et al.*, 1999; Shera and Guinan, 1999). However it has been suggested that the presence of a low-frequency bias tone can suppress the component generated at the characteristic region of the DPOAE (Talmadge *et al.*, 1999; Bian and Scherrer, 2007), simplifying the regions of the cochlea contributing to the DPOAEs. Presumably, the OP_{Bias} derived from DPOAEs is primarily a measure of the OC position at the overlapping region and would be dependent on the probe frequency, which could be varied to investigate the resting position of the OC at different regions throughout the cochlea.

B. The OP_{Boltz} is a measure of local transduction

One motive for simultaneously measuring the OP_{Boltz} derived from ST and SM at the same region along the length of the cochlea in the present study was to demonstrate that Boltzmann analysis of CM is an accurate measure of local transduction. It had recently been suggested that distortion products in the CM were generated at distal regions within the cochlea, particularly for low-level stimuli (Zheng *et al.*, 2008). For stimulus levels that permitted reliable Boltzmann fits (above 80 dB SPL, Fig. 5), our OP_{Boltz} measurements derived from ST and SM CM were modulated in-phase and by a similar amount by the low-frequency bias, even though the ST and SM CM amplitudes differed by a factor of around 10, and the CM waveforms were of opposite polarity. If electrical distortion components from CM generated at more apical regions in the cochlea summated with CM generated in the basal turn where our measurements of ST and SM CM were made, our ST and SM OP_{Boltz} measurements would have been modulated 180 degrees out-of-phase during low-frequency biasing because remote potentials spread electrotonically to the basal turn would not invert polarity across the cochlear partition, and our analysis incorporated the inversion of the locally generated CM. Our data suggest that distortion in the CM is generated at the cochlear partition and that Boltzmann analysis of this CM is an accurate measure of local transduction, at least for 500 Hz stimulus levels above 80 dB SPL.

Similar changes in the amplitude and phase of the distortion products have been observed using the whole-cell voltage clamp technique to monitor the receptor current of OHCs (Takahashi and Santos-Sacchi, 1999), and Olson (2004) demonstrated that distortion products in fluid pressure measurements made close to the basilar membrane were generated locally, rather than at regions distal to the recording location. These studies provide further evidence that the dominant nonlinearity underlying distortion generation in the CM is produced by *local* OHC electromotility and that the distortion products in the CM can be used to estimate the OP of local OHC transduction.

C. Cochlear distortion products behave as predicted from a first-order Boltzmann function

Similar to results from previous studies (Kirk *et al.*, 1997; Sirjani *et al.*, 2004), the present study found that the amplitudes of the even-order distortion products from the CM and ear canal were minimal when the OP measure (OP_{Boltz} or OP_{Bias}) was near zero and maximal when the OP measure was displaced from zero. Conversely, the odd-order distortion products were maximal when the OP was near zero and minimal when the OP was displaced by $\frac{1}{4}$ to $\frac{1}{5}$ the peak-to-peak pressure of the probe. These changes suggest that the transducer generating distortion in the cochlea can be modeled with a first-order Boltzmann distribution function. However, there were several differences between the modeled and measured distortion products that suggest that slightly more accurate assessment of cochlear transduction might be achieved with more complex analysis methods, as discussed previously in detail by Patuzzi and Moleirinho (1998) with regard to Boltzmann analysis of the CM.

With the OP of a first-order transducer at zero, there should have been no even-order distortion products generated; however, often the even-order distortion products in the CM and DPOAE were not reduced to zero when the OP measure was near zero [Fig. 6(C)]. Furthermore, the relative phase of the even-order distortion products generated by such a transducer should have been either 90° or 270° (depending on the polarity of the OP and with the phase of f1 at zero), although often their phases were as much as 15° different from this when the OP was maximally displaced by the bias tone and as much as 90° different when the OP was near zero [Fig. 6(D)]. One likely explanation for the differences between the modeled and the measured distortion products is that there were additional non-dominant sources of distortion. Another possible explanation is that the CM and DPOAE distortion products are generated by a region of the cochlea along which the transducer properties change slightly, and therefore the distortion products are a slightly complex weighted sum of the contributing region, producing differences between our measured data and our estimates based on a single transducer.

In the case of the CM recorded from the base of the cochlea, which is predominantly generated by local hair cells (Dallos *et al.*, 1972; Patuzzi *et al.*, 1989), the other sources might have been either (1) other sources of nonlinearity in the local transduction, which also produced distortion of the receptor current but only dominated when the OP of the foremost nonlinearity was near zero, (2) distortion products generated by OHC at other regions of the cochlea, which spread electrotonically to the recording location and had different characteristics (phase and amplitude) than those generated by basal hair cells, (3) distorted OC vibration generated at other regions in the cochlea, which then produced additional vibration at the base of the cochlea, or (4) low-level distortion of the acoustic stimulus resulting in a non-sinusoidal stimulus. Whatever the other source(s) of distortion in the CM were, it resulted in only minor deviations from the values predicted by a first-order Boltzmann equation.

In the case of the DPOAEs, many previous studies have

researched the sources of DPOAEs in the cochlea, with the general consensus that there are two main contributors, a linear reflection component and a nonlinear distortion component (Talmadge *et al.*, 1999; Shera and Guinan, 1999). Multiple contributions to the DPOAEs might result in complex modulations of the DPOAE amplitudes throughout the bias that could not be modeled as a single transducer with a Boltzmann distribution. However, in the present study the DPOAEs often followed a bimodal modulation, suggesting either that one of the components typically dominated the DPOAE generation or that both components contributed equally, producing similar DPOAE modulations. We did not attempt to investigate the contribution from either component, and most likely there was some complex summation of the two sources, which decreased the accuracy of our OP_{Bias} measure slightly. Given the similarity between the OP_{Boltz} measure from basal turn CM, which is a more local measure and the OP_{Bias} from the DPOAEs measured in the present study, this inaccuracy is most likely small.

It is worth noting here that in the present study our Boltzmann function expressed OP_{Boltz} in terms of a static pressure in the ear canal [using $S^*P^*_{sat}(\text{input}+\text{OP})$ in the equation], rather than as a measure unrelated to ear canal pressure but still proportional to the asymmetry of transduction [using $(S^*P_{sat})+\text{OP}$ in the equation]. Patuzzi and Moleirinho (1998) used the latter method as they assumed that some OP changes might have been due to slight micromechanical changes in the position of the OC, unrelated to the ear canal pressure. While there are several situations that might lead to a difference between the OP estimated using these different methods, such as changes in middle ear conductance or a change in the OC compliance during a low-frequency bias, we routinely compared our OP measure to that obtained using the method employed by Patuzzi and Moleirinho (1998) and found the differences to be minimal, only differing substantially for changes in probe level, for which it was not possible to determine the preferable method as OP changes with stimulus level might have been produced by OC compliance changes *or* micromechanical changes in the OC position.

D. Probe and bias level effects on OP estimate

Ideally, we would want to use probe levels that generate distortion products in the ear canal or CM that are well above the noise floor of the recording and that generate CM waveforms that are sufficiently saturated to allow accurate Boltzmann analysis. However, relatively high probe levels (e.g., above 100 dB SPL for 500 Hz tones) often caused unwanted adaptation of the OP measure [Fig. 5(C)] and might lead to temporary threshold shifts particularly in the case of high-frequency probe tones (at levels above approximately 85 dB SPL for continuous stimulation; Chang and Norton, 1996). Therefore, the ideal range of probe levels used for the present techniques is most likely limited to around 65–80 dB SPL for high-frequency probe tones (i.e., around 4 kHz) and around 80–95 dB SPL for low-frequency tones (i.e., around 500 Hz). These probe levels did not alter the CAP thresholds or EP measurements before or after presentation of the probes, suggesting that they did not cause any temporary or permanent changes in cochlear function the guinea pigs used in this study, other than a rapid slight adaptation which occurred at the immediate onset of the probe with a time constant of around 1 s (which was not included in our analysis of the responses), similar to that observed by Kujawa and Liberman (2001), and which was most likely related to efferent activity. As stimulus levels were measured close to the tympanic membrane, they were effectively 5–10 dB lower than if they had been presented and calibrated free field since no ear canal gain is available in this preparation.

Such levels are on the upper limit of what would be acceptable for human DPOAE testing (although levels as high as 85 dB SPL at 4 kHz have been used; Lonsbury-Martin *et al.*, 1990). We are currently evaluating whether OP_{Bias} can be determined from DPOAEs that does not require the transducer to be saturated.

While the slope of the relationship between the amplitude of the even-order distortion products and the OP changed over a range of sound levels (Fig. 5), this did not affect our estimation of the OP_{Bias} because it was not based solely on the amplitude of the distortion products. Rather, it was based on notches in the amplitude of the modulated distortion products during the bias, as outlined in Sec. II F, and the overall ratio of the distortion product to f1 (i.e., the *y*-axis of our *V*-plots) was not a key factor in our analysis.

Generally, different bias tone levels (from 95 to 115 dB SPL) simply produced different levels of OP modulation (Fig. 8), producing only a slight change in the average OP measure [Fig. 7(A)], which most likely reflects a small amount of OP adaptation to the bias tone. It was often necessary to use bias levels of 110 dB SPL in order to produce bimodal modulation of the even-order distortion products and a modulation of the odd-order distortion products that allowed estimation of notches in its amplitude, which was a limitation of the technique used to estimate OP_{Bias}. However, it is important to note that the cochlea is generally insensitive to such low-frequency stimuli due to the high-pass filtering characteristics of the helicotrema, as described below, and following exposure to the highest level bias tones used in the present study, cochlea sensitivity was unchanged. While the presence of the bias might have been responsible for the slight changes in the OP measure over the range of bias levels, these changes were negligible in comparison to the experimentally induced changes in OP such as that produced by hypoxia.

E. Referencing OP_{Bias} to the probe level

In order to compare the OP_{Boltz} and OP_{Bias} measures, it was necessary to reference the OP_{Bias} values to the level of the probe stimulus in the ear canal using characteristic changes in the distortion products as outlined in Sec. II F of the present study. This allowed estimation of the relative displacement of the OP at the region where the DPOAEs were generated in the cochlea, and it overcame the problem that the level of the distortion products in the ear canal could

change due to factors unrelated to the OP, such as conductive losses as the DPOAEs were transmitted from the cochlea to the ear canal.

Similar to results published in previous studies (Takahashi and Santos-Sacchi, 1999; Bian and Scherrer, 2007), we found amplitude minima or "notches" in the odd-order distortion products in the CM and DPOAEs during modulations of the OP. From our model of distortion generated by a first-order Boltzmann function, we found that for inputs which sufficiently saturated the transducer, these notches occurred at certain OP values that were related to the probe level. With the assumption that the modulated odd-order distortion products adhered to the relationship between the OP and notches in the modulated odd-order distortion products, we were able to relate the OP_{Bias} values to the probe based on the notches in the modulated odd-order distortion products. The values of the OP_{Boltz} and the corrected OP_{Bias} derived from the CM were similar [Fig. 4(F)], as was the magnitude of the overall changes in the OP_{Bias} and OP_{Boltz} during induced hypoxias, suggesting that the method for relating the OP_{Bias} to the probe level in the ear canal was viable.

The finding that the OP_{Bias} derived from the DPOAEs was modulated two to three times more than the OP_{Bias} derived from the basal turn CM during the low-frequency bias (Fig. 9) suggested that the bias produced a larger displacement of the OC near the 4–4.8 kHz characteristic frequency region than at the base of the cochlea. Presumably, this reflects the mechanical impedance change to low frequencies along the length of the cochlea (von Békésy, 1960; Patuzzi, 1996). Previous studies have used low-frequency modulation of odd-order DPOAEs to non-invasively estimate the sensitivity of the OC displacement to a low-frequency bias tone in humans (Bian and Scherrer, 2007; Marquardt *et al.*, 2007), changing the bias tone frequency between 15 and 480 Hz and examining the differences in the notches of the odd-order distortion products.

It is worth noting here that the same technique might be used to examine sensitivity of the OC displacement to the high-frequency probe tone(s) because the notches in the odd-order DPOAEs are a function of the sensitivity to both the bias tone *and* the probe tone(s) (i.e., notches in the biased odd-order distortions occur at $\frac{1}{4}$ or $\frac{1}{5}$ the pressure of the probe tone, which depends on the sensitivity to the probe tone). Interestingly, this might provide a non-invasive objective measure of the OC displacement and sensitivity throughout the entire cochlea (i.e., if the bias frequency is held constant while the probe frequencies are swept).

F. Estimating the transducer OP

This study did not attempt to determine the dominant nonlinearity underlying distortion in the ear canal and cochlea, be it nonlinearity of MET channel compliance (Howard and Hudspeth, 1988; Jaramillo *et al.*, 1993), nonlinearity of MET channel conductance (Corey and Hudspeth, 1979) that produces mechanical distortion via reverse-transduction (Kirk and Yates, 1998), or nonlinearity in the active process itself, whatever that process may be (see Hudspeth, 2008 for recent review). Rather, we suggest that the dominant nonlinearity underlying distortion in both the ear canal and the CM is similar and that this distortion is related to the displacement of the OC.

Because we did not have a direct measure of the OC displacement, we estimated the OP as a measure that was *equivalent* to displacements of the OC produced by an acoustic stimulus in the ear canal. Measuring the OP relative to the probe stimulus allowed us to compare the OC displacement produced by the high-frequency probe tones (500 Hz or 4 and 4.8 kHz) to that produced by the low-frequency bias tone (4.8 Hz).

The 115 dB SPL, 4.8 Hz tone was equivalent to 32 Pa peak-to-peak pressure in the ear canal but, relative to the 500 Hz, 90 dB SPL probe tone (which was 1.79 Pa peak-to-peak), only produced a 0.19 Pa modulation of the OP_{Boltz} (Figs. 7 and 8). This suggests that the 4.8 Hz tone was 45 dB less effective in displacing the OC than the 500 Hz tone. This difference is most likely due to the acoustic high-pass filtering of the helicotrema, which attenuates frequencies lower than 100 Hz (Dallos, 1970; Ruggero *et al.*, 1986; Marquardt *et al.*, 2007). Presumably, higher-frequency bias tones would have produced a relatively larger displacement of the OC and greater modulation of the OP and would explain the more sensitive modulation of the DPOAEs reported by Bian and Scherrer (2007) who used 25–100 Hz bias tones.

We have recent evidence that suggests that the sensitivity differences between the high-frequency and bias tones are abnormal with auditory disorders such as endolymphatic hydrops (Marquardt *et al.*, 2007; Hensel *et al.*, 2007). If the helicotrema were partially or fully blocked off due to distension of Reissner's membrane with endolymphatic hydrops, we might expect to see a larger sensitivity to the low-frequency bias tone due to the reduced effectiveness of the helicotrema as a low-frequency shunt. This has been confirmed experimentally in recent experiments where gel was injected into the cochlear apex (Salt *et al.*, 2009).

Finally, low-frequency biasing of DPOAEs might provide a twofold indication of hydrops. First, if endolymphatic hydrops generates an over-pressure in SM, displacing the OC toward ST, the OP_{Bias} estimate obtained using low-frequency DPOAE biasing would be abnormally positive. Second, if endolymphatic hydrops resulted in distension of Reissner's membrane and blockage of the helicotrema, we might expect to find an abnormally high sensitivity to the low-frequency bias tone, as evident in the level of the bias required to produce notches in the odd-order distortion product.

ACKNOWLEDGMENTS

This study was supported by research Grant No. RO1 DC01368 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health and by a grant to D.J.B. from the Ménière's Research Fund Group of New South Wales, Australia.

Bian, L. (**2004**). "Cochlear compression: Effects of low-frequency biasing on quadratic distortion product otoacoustic emission," J. Acoust. Soc. Am. **116**, 3559–3571.

Bian, L. (**2006**). "Spectral fine-structures of low-frequency modulated distortion product otoacoustic emissions," J. Acoust. Soc. Am. **119**, 3872–3885.

Bian, L., Chertoff, M. E., and Miller, E. (**2002**). "Deriving a cochlear transducer function from low-frequency modulation of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **112**, 198–210.
Bian, L., and Scherrer, N. M. (**2007**). "Low-frequency modulation of distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **122**, 1681–1692.
Brownell, W. E., Bader, C. R., Bertrand, D., and de Ribaupierre, Y. (**1985**). "Evoked mechanical responses of isolated cochlear outer hair cells," Science **227**, 194–196.
Chang, K. W., and Norton, S. J. (**1996**). "The effects of continuous versus interrupted noise exposures on distortion product otoacoustic emissions in guinea pigs," Hear. Res. **96**, 1–12.
Cheatham, M. A., and Dallos, P. (**1994**). "Stimulus biasing: A comparison between cochlear hair cell and organ of Corti response patterns," Hear. Res. **75**, 103–113.
Corey, D. P., and Hudspeth, A. J. (**1979**). "Ionic basis of the receptor potential in a vertebrate hair cell," Nature (London) **281**, 675–677.
Corey, D. P., and Hudspeth, A. J. (**1983**). "Analysis of the microphonic potential of the bullfrog's sacculus," J. Neurosci. **3**, 942–961.
Crawford, A. C., Evans, M. G., and Fettiplace, R. (**1989**). "Activation and adaptation of transducer currents in turtle hair cells," J. Physiol. (London) **419**, 405–434.
Dallos, P., Billone, M. C., Durrant, J. D., Wang, C., and Raynor, S., (**1972**). "Cochlear inner and outer hair cells: Functional differences," Science **177**, 356–358.
Dallos, P. (**1970**). "Low-frequency auditory characteristics: Species dependence," J. Acoust. Soc. Am. **48**, 489–499.
Dallos, P. (**1992**). "The active cochlea," J. Neurosci. **12**, 4575–4585.
Davis, H. (**1983**). "An active process in cochlear mechanics," Hear. Res. **9**, 79–90.
Evans, B. N., and Dallos, P. (**1993**). "Stereocilia displacement induced somatic motility of cochlear outer hair cells," Proc. Natl. Acad. Sci. U.S.A. **90**, 8347–8351.
Frank, G., and Kössl, M. (**1996**). "The acoustic two-tone distortions 2f1-f2 and f2-f1 and their possible relation to changes in the operating point of the cochlear amplifier," Hear. Res. **98**, 104–115.
Garner, C. A., Neely, S. T., and Gorga, M. P. (**2008**). "Sources of variability in distortion product otoacoustic emissions," J. Acoust. Soc. Am. **124**, 1054–1067.
Halsey, K., Skjönsberg, A., Ulfendahl, M., and Dolan, D. F. (**2005**). "Efferent-mediated adaptation of the DPOAE as a predictor of aminoglycoside toxicity," Hear. Res. **201**, 99–108.
Hensel, J., Scholz, G., Hurttig, U., Mrowinski, D., and Janssen, T. (**2007**). "Impact of infrasound on the human cochlea," Hear. Res. **233**, 67–76.
Hirsh, I. J., and Ward, W. D. (**1952**). "Recovery of the acoustic threshold after strong acoustic stimulation," J. Acoust. Soc. Am. **24**, 131–141.
Holton, T., and Hudspeth, A. J. (**1986**). "The transduction channel of hair cells from the bull-frog characterized by noise analysis," J. Physiol. **375**, 195–227.
Howard, J., and Hudspeth, A. J. (**1988**). "Compliance of the hair bundle associated with gating of MET channels in the bullfrog's saccular hair cell," Neuron **1**, 189–199.
Hudspeth, A. J. (**2008**). "Making an effort to listen: Mechanical amplification in the ear," Neuron **59**, 530–545.
Jaramillo, F., Markin, V. S., and Hudspeth, A. J. (**1993**). "Auditory illusions and the single hair cell," Nature (London) **364**, 527–529.
Kemp, D. T. (**1986**). "Otoacoustic emissions, travelling waves and cochlear mechanisms," Hear. Res. **22**, 95–104.
Kim, D. O. (**1986**). "Active and nonlinear cochlear biomechanics and the role of outer-hair-cell subsystem in the mammalian auditory system," Hear. Res. **22**, 105–114.
Kirk, D. L., Moleirinho, A., and Patuzzi, R. B. (**1997**). "Microphonic and DPOAE measurements suggest a micromechanical mechanism for the 'bounce' phenomenon following low-frequency tones," Hear. Res. **112**, 69–86.
Kirk, D. L., and Patuzzi, R. B. (**1997**). "Transient changes in cochlear potentials and DPOAEs after low-frequency tones: The 'two-minute bounce' revisited," Hear. Res. **112**, 49–68.
Kirk, D. L., and Yates, G. K. (**1998**). "Enhancement of electrically evoked oto-acoustic emissions associated with low-frequency stimulus bias of the basilar membrane towards scala vestibuli," J. Acoust. Soc. Am. **104**, 1544–1554.
Kossowski, M., Mom, T., Guitton, M., Poncet, J. L., Bonfils, P., and Avan, P. (**2001**). "Fine alterations of distortion-product otoacoustic emissions after moderate acoustic overexposure in guinea pigs," Audiology **40**, 113–122.
Kros, C. J. (**1996**). "Physiology of mammalian cochlear hair cells," in *The Cochlea*, Springer Handbook of Auditory Research Vol. **8**, edited by P. Dallos, R. R. Fay, and A. N. Popper (Springer US, NY), pp. 318–385.
Kujawa, S. G., and Liberman, M. C. (**2001**). "Effects of olivocochlear feedback on distortion product otoacoustic emissions in guinea pig," J. Assoc. Res. Otolaryngol. **2**, 268–278.
Liberman, M. C., Zuo, J., and Guinan, J. J., Jr. (**2004**). "Otoacoustic emissions without somatic motility: Can stereocilia mechanics drive the mammalian cochlea?" J. Acoust. Soc. Am. **116**, 1649–1655.
Lonsbury-Martin, B. L., Harris, F. P., Stagner, B. B., Hawkins, M. D., and Martin, G. K. (**1990**). "Distortion product emissions in humans. I. Basic properties in normally hearing subjects," Ann. Otol. Rhinol. Laryngol. Suppl. **147**, 3–14.
Lukashkin, A. N., and Russell, I. J. (**1997**). "The voltage dependence of the mechanoelectrical transducer modifies low frequency outer hair cell electromotility in vitro," Hear. Res. **113**, 133–139.
Lukashkin, A. N., and Russell, I. J. (**1998**). "A descriptive model of the receptor potential nonlinearities generated by the hair cell mechanoelectrical transducer," J. Acoust. Soc. Am. **103**, 973–980.
Lukashkin, A. N., and Russell, I. J. (**2002**). "Modifications of a single saturating non-linearity account for post-onset changes in 2f1-f2 distortion product otoacoustic emission," J. Acoust. Soc. Am. **112**, 1561–1568.
Marquardt, T., Hensel, J., Mrowinski, D., and Scholz, G., (**2007**). "Low-frequency characteristics of human and guinea pig cochleae," J. Acoust. Soc. Am. **121**, 3628–3638.
Mrowinski, D., Scholz, G., Krompass, S., and Nubel, K., (**1996**). "Diagnosis of endolymphatic hydrops by low-frequency masking," Audiol. Neuro-Otol. **1**, 125–134.
Nieder, P., and Nieder, I. (**1968a**). "Some effects of tonal interactions as seen in the cochlear microphonic," J. Acoust. Soc. Am. **43**, 1092–1106.
Nieder, P., and Nieder, I. (**1968b**). "Studies of two-tone interaction as seen in the guinea pig microphonic," J. Acoust. Soc. Am. **44**, 1409–1422.
Olson, E. S. (**2004**). "Harmonic distortion in intracochlear pressure and its analysis to explore the cochlear amplifier," J. Acoust. Soc. Am. **115**, 1230–1241.
Olzowy, B., von Gleichenstein, G., Canis, M., Plesnila, N., and Mees, K. (**2008**). "Complex level alterations of the 2f (1)-f (2) distortion product due to hypoxia in the guinea pig," Eur. Arch. Otorhinolaryngol. (in press).
Patuzzi, R., and Moleirinho, A. (**1998**). "Automatic monitoring of mechano-electrical transduction in the guinea pig cochlea," Hear. Res. **125**, 1–16.
Patuzzi, R., and Rajan, R. (**1990**). "Does electrical stimulation of the crossed olivo-cochlear bundle produce movement of the organ of Corti?" Hear. Res. **45**, 15–32.
Patuzzi, R., and Sellick, P. M. (**1983**). "A comparison between basilar membrane and inner hair cell receptor potential input-output functions in the guinea pig cochlea," J. Acoust. Soc. Am. **74**, 1734–1741.
Patuzzi, R., Sellick, P. M., and Johnstone, B. M. (**1984a**). "The modulation of the sensitivity of the mammalian cochlea by low frequency tones. I. Primary afferent activity," Hear. Res. **13**, 1–8.
Patuzzi, R., Sellick, P. M., and Johnstone, B. M. (**1984b**). "The modulation of the sensitivity of the mammalian cochlea by low frequency tones. III. Basilar membrane motion," Hear. Res. **13**, 19–27.
Patuzzi, R. B. (**1996**). "Cochlear micromechanics and macromechanics," in *The Cochlea*, Springer Handbook of Auditory Research Vol. **8**, edited by P. Dallos, R. R. Fay, and A. N. Popper (Springer US, NY), pp. 186–257.
Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989**). "The origin of the low-frequency microphonic in the first cochlear turn of guinea-pig," Hear. Res. **39**, 177–188.
Rebillard, G., and Lavigne-Rebillard, M. (**1992**). "Effect of reversible hypoxia on the compared time courses of endocochlear potential and 2f1-f2 distortion products," Hear. Res. **62**, 142–148.
Ruggero, M. A., Robles, L., and Rich, N. C. (**1986**). "Basilar membrane mechanics at the base of the chinchilla cochlea. II. Responses to low-frequency tones and relationship to microphonics and spike initiation in the VIII nerve," J. Acoust. Soc. Am. **80**, 1375–1383.
Russell, I. J., and Kossl, M. (**1992**). "Modulation of hair cell voltage responses to tones by low-frequency biasing of the basilar membrane in the guinea pig cochlea," J. Neurosci. **12**, 1587–1601.
Salt, A. N., and Demott, J. E. (**1999**). "Longitudinal endolymph movements and endocochlear potential changes induced by stimulation at infrasonic frequencies," J. Acoust. Soc. Am. **106**, 847–856.
Salt, A. N., Brown, D. J., Hartsock, J. J., and Plontke, S. K. (**2009**). "Sus-

tained displacements of the organ of Corti and hypersensitivity to infrasound induced by gel injections into the cochlear apex," Hear. Res. (to be published).
Salt, A. N., Hensen, H. E., Gill, R. M., and Sirjani, D. B. (**2005**). "Even-order distortion as an indicator of endolymphatic hydrops," in Proceedings of the Fifth International Symposium on Meniere's Disease and Inner Ear Homeostasis Disorders, The House Institute, Los Angeles, CA, edited by D. J. Lim.
Santos-Sacchi, J. (**1991**). "Reversible inhibition of voltage-dependent outer hair cell motility and capacitance," J. Neurosci. **11**, 3096–3110.
Shera, C. A., and Guinan, J. J., Jr. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Sinyor, A., and Laszlo, C. A. (**1973**). "Acoustic behavior of the outer ear of the guinea pig and the influence of the middle ear," J. Acoust. Soc. Am. **54**, 916–921.
Sirjani, D. B., Salt, A. N., Gill, R. M., and Hale, S. A. (**2004**). "The influence of transducer operating point on distortion generation in the cochlea," J. Acoust. Soc. Am. **115**, 1219–1229.
Strelioff, D. (**1973**). "A computer simulation of the generation and distribution of cochlear potentials," J. Acoust. Soc. Am. **54**, 620–629.
Takahashi, S., and Santos-Sacchi, J. (**1999**). "Distortion component analysis of outer hair cell motility-related gating charge," J. Membr. Biol. **169**, 199–207.
Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **105**, 275–292.
Valk, W. L., Wit, H. P., and Albers, F. W. (**2004**). "Evaluation of cochlear function in an acute endolymphatic hydrops model in the guinea pig by measuring low-level DPOAEs," Hear. Res. **192**, 47–56.
von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).
Weiss, T. F., and Leong, R. (**1985**). "A model for signal transmission in an ear having hair cells with free-standing stereocilia. IV. Mechanoelectric transduction stage," Hear. Res. **20**, 175–195.
Xenellis, J. E., Linthicum, F. H., Jr., Webster, P., and Lopez, R. (**2004**). "Basilar membrane displacement related to endolymphatic sac volume," Laryngoscope **114**, 1953–1959.
Yates, G. K., and Kirk, D. L. (**1998**). "Cochlear electrically evoked emissions modulated by mechanical transduction channels," J. Neurosci. **18**, 1996–2003.
Zenner, H. P. (**1986**). "Motile responses in outer hair cells," Hear. Res. **22**, 83–90.
Zheng, J., Ren, T., and Nuttall, A. (**2008**). "Exploring the generation of low frequency cochlear microphonic distortion," Poster Presented at the 31st Midwinter ARO Meeting, AZ, Abstract No. 1119.

Inverse-solution method for a class of non-classical cochlear models

Egbert de Boer
Academic Medical Center, University of Amsterdam, Room D2-226, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands; Oregon Hearing Research Center, NRC04, Oregon Health & Science University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239-3098

Alfred L. Nuttall[a)]
Oregon Hearing Research Center, NRC04, Oregon Health & Science University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239-3098; Kresge Hearing Research Institute, University of Michigan, 1301 E. Ann Street, Ann Arbor, Michigan 48109-0506

(Received 29 October 2008; revised 27 January 2009; accepted 28 January 2009)

Measurements of distortion-product (DP) waves inside the cochlea have led to a conception of wave propagation that is at variance with the "classical" attitude. Of the several alternatives that have been proposed to remedy this situation, the feed-forward model could be a promising one. This paper describes a method to apply the inverse solution with the aim to attain a feed-forward model that accurately reproduces a measured response. It is demonstrated that the computation method is highly successful. Subsequently, it is shown that in a feed-forward model a DP wave generated by a two-tone stimulus is almost exclusively a forward-traveling wave which property agrees with the nature of the experimental findings. However, the amplitude of the computed DP wave is only substantial in the region where the stimulation patterns of the two primary tones overlap. In addition, the model developed cannot explain coherent reflection for single tones. It has been suggested that a forward transversal DP wave induced by a (retrograde) compression wave could be involved in DP wave generation. This topic is critically evaluated.
 [DOI: 10.1121/1.3083240]

PACS number(s): 43.64.Kc, 43.64.Bt [BLM] Pages: 2146–2154

I. INTRODUCTION

Models of cochlear mechanics usually consist of a dual fluid-filled channel of which the parts are separated by a flexible partition. The solution of the model equation then represents the equilibrium between the hydrodynamics of the fluid and the dynamics of the partition—which is dominated by the basilar membrane (BM) and the cells of the organ of Corti. In a "classical" model, the mechanical properties of the cochlear partition depend only on the *local* longitudinal coordinate. Classical models have abundantly been used for a great variety of purposes. "Non-classical" models contain mechanical properties that depend on variables at more than one location. Such models have been used to explain properties that could not be explained by a classical model. In the field of non-classical models, feed-forward models have been considered in more detail (a few examples: Steele *et al.*, 1993; Geisler and Sang, 1995; Wen and Boahen, 2003). In a feed-forward model, the pressure at location x depends on the velocity at the same location (the "passive" or classical component of the dynamics) as well as on the velocity at location $x-\Delta x$ (the non-classical component) where Δx is a positive constant—generally taken to be equal to several times the diameter of an outer hair cell (OHC). This type of model has unique properties in its modes of wave propagation (de Boer, 2007) which makes it suitable to explain wave-propagation properties of observed distortion products (DPs) (for specific references, see further on). To clarify why this would be important, a few descriptive steps have to be taken.

Step 1: A classical model. The response of the viable, living cochlea is quite different from that of the *postmortem* cochlea. This implies that the dynamics of the cochlear partition is quite different in these two cases. In the *postmortem* cochlea, the BM has passive properties; its dynamics can be described by an impedance which has stiffness and mass, and a positive resistance. In the living cochlea, an additional mechanism causes the response to show a pronounced peak, and when we assume linearity, this effect can be described by an additional component, the "active" component, of the BM impedance. It is generally agreed that the effects of this additional impedance are caused by OHCs; the precise manner in which these cells perform their task is greatly unknown at present. One factor involved may be the mechanical action of the Internal Spiral Sulcus, as described in the literature, see de Boer (1993) and Steele and Lim (1999). A possible mechanism taking care of the necessary frequency selectivity may reside in resonance of the tectorial membrane, see, e.g., Neely and Kim (1986) for a model based on ideas by Allen (1980) and Zwislocki and Kletsky (1979). de Boer (1996), Patuzzi (1996), and Robles and Ruggero (2001) have written reviews. It is repeated that in this context the cochlear model, including the action of the OHCs, is assumed linear.

Step 2. When the cochlea is stimulated by two primary

a)Author to whom correspondence should be addressed. Electronic mail: nuttall@ohsu.edu

tones, with frequencies f_1 and f_1, the OHCs, being intrinsically nonlinear physiological transducers, will generate harmonic components with the same frequencies and their multiples, but also intermodulation components, which we will term DPs. These DPs are mainly generated in the *overlap region* which is the region where the excitation patterns of the two primary tones overlap. One of the most important of those DPs has the frequency $2f_1-f_2$. From the overlap region, two waves with the DP frequency are assumed to originate. One travels toward the helicotrema, and the other one toward the stapes. Physically, these two waves should have the same or closely related propagation properties. However, in experiments, they manifest themselves quite differently. The apical DP wave is easily observable but the reverse DP wave is not. In fact, where one would expect that reverse wave to exist, between the overlap region and the stapes, it is impossible or hard to find. In that region a forward wave dominates (Ren, 2004; He *et al.*, 2007, 2008; de Boer *et al.*, 2008; Dong and Olson, 2008). This curious phenomenon forms a challenge to theorists.

Step 3. One of the theories advanced to explain this property involves feed-forward. As has been observed (Steele *et al.*, 1993) and proven rigorously (de Boer, 2007), the mechanism of feed-forward, when adjusted to show wave amplification for forward-traveling waves, demonstrates wave attenuation for reverse waves. That is, the propagation of a reverse wave differs from that of a forward wave. Finding out the main characteristics of cochlear models with feed-forward (and feed-backward) with respect to measured data forms the main topic of the present paper.

Step 4. Another explanation of the anomalous direction of wave propagation of DP waves has been brought forward; it involves compression waves (Ren, 2004). Briefly, the proposed mechanism is as follows. In the region of overlap, the OHCs form a large number of tiny sound sources oscillating with the DP frequency. Together all these sources give rise to a compression wave in the fluid. Ideally, that wave has an infinite velocity of propagation, and it invades all tissues and spaces of the inner ear instantly. The mechanism by which a compression wave turns into a transversal wave is the same as that by which we hear bone-conducted sound. The compression wave reaches the region of the oval and round windows. Because the mechanical impedances of the two membranes are quite different, the two membranes will start to oscillate with different velocities. As a result, a "normal" propagating wave starts at the windows, having the DP frequency as its frequency. A short discussion of compression waves is included in this paper.

In earlier work of the present authors, the *inverse-solution method* (de Boer, 1995a, 1995b; de Boer and Nuttall, 1999) has been used for several purposes (e.g., to explain intensity effects: de Boer and Nuttall, 2000; to assess tones versus noise stimuli: de Boer and Nuttall, 2002; to analyze spontaneous oscillations of the BM: de Boer and Nuttall, 2006b; to analyze general wave-propagation properties: de Boer *et al.* 2007). In the inverse solution, the mechanical impedance of the BM is derived from a given BM response curve. That work was always centered at a classical cochlear model.

In the present study, it is shown how the inverse method can be extended to a feed-forward model. This application will result in a non-classical model, a feed-forward model, which is shown to provide an accurate reproduction of the response actually measured in a physiological experiment. An application in a different sense is illustrated for the case of a DP wave generated by nonlinear distortion of OHCs. This extension forms the main theme of the present work. Discussion of feed-backward and combination of feed-forward with feed-backward is beyond the scope of this paper.

II. DERIVATION: THE INVERSE SOLUTION

We will first consider the case of the classical, linear model, with two fluid-filled channels. In such a model, the local velocity of the BM only depends on the local pressure difference across it, and not on influences from elsewhere, from other locations. In that case the mechanics of the organ of Corti is lumped into the concept of a mechanical impedance (a point-impedance), called the BM impedance Z_{BM}. This impedance relates (in a way to be detailed further on) the pressure across the BM at a point x along the length of the BM to the velocity of the BM at the *same* location x. That impedance, obviously a function of location x, is converted into a diagonal matrix $\mathbf{Z}_{\mathrm{BM}}$. In this step we have tacitly replaced the continuous variable x by a finite sequence of N samples. The model equation then reads (de Boer *et al.*, 2007):

$$\left(i\omega\rho\mathbf{G}+\tfrac{1}{2}\mathbf{Z}_{\mathrm{BM}}\right)\mathbf{v}=-i\omega\rho\mathbf{S}v_{\mathrm{st}}. \tag{1}$$

The matrix $\mathbf{G}$ of size $N\times N$ (which constitutes the Green's function) represents the hydrodynamics of the fluid inside the model. The vector $\mathbf{S}$ (a column vector) is the stapes propagator (see Mammano and Nobili, 1993) and v_{st} is the stapes velocity. The "stylized" model of the cochlea has the form of a rectangular block in which the BM occupies a fraction (in our case 0.2) of the width, as in earlier work (de Boer, 1995a, 1995b). From Eq. (1) the BM velocity $\mathbf{v}$ (again a column vector) can be solved for a given form of $\mathbf{Z}_{\mathrm{BM}}$—this constitutes the *forward* solution.

The normal procedure of the *inverse* solution consists of two parts. First, from the given BM velocity $\mathbf{v}$ the pressure $\mathbf{p}$ that is solely due to hydrodynamics is determined. Normally, this is formed from the combination of the first term and the right-hand member of Eq. (1):

$$\mathbf{p}=i\omega\rho\mathbf{G}\mathbf{v}+i\omega\rho\mathbf{S}v_{\mathrm{st}}. \tag{2}$$

It is important to notice that the pressure $\mathbf{p}$ is solely determined by the hydrodynamics of the fluid because the matrix $\mathbf{G}$ and the vector $\mathbf{S}$ solely depend on the geometry of the bony shell of the cochlea. In the second step the effective BM impedance Z_{eff} is determined from the so-computed pressure $\mathbf{p}$ and the (given) BM velocity $\mathbf{v}$ (all variables are tacitly considered to be functions of location x):[1]

$$Z_{\text{eff}} = -2p/v. \tag{3}$$

In fact, Z_{eff} is a function of location x, but for reasons of clarity and simplicity we will omit the dependence on x from our equations wherever that is feasible.

To verify the accuracy of the procedure, the *resynthesized response* of the model is computed by substituting the derived impedance Z_{eff}, converted into a diagonal matrix $\mathbf{Z}_{\text{eff}}$, into the model Eq. (1) and solving that equation for $\mathbf{v}$. Because an inverse solution is known to be like an "ill-posed problem," this step is not superfluous (for more details on resynthesis see de Boer and Nuttall, 1999).

It should now be recognized that the so-derived BM impedance Z_{eff} is composed of a passive part that represents the impedance of the BM in the *postmortem* cochlea and an active part that induces the response of the viable cochlea to deviate so drastically from the *postmortem* response. This distinction is of prime importance.

In a *feed-forward* model, the OHC contribution from the BM velocity at location x results in an added pressure at location $(x+\Delta x)$, where Δx is positive. Assume, for a moment, that this is the only effect of the dynamics of the BM; this means that we temporarily omit the mechanical dynamics of the (passive) BM. Introduce X_{ffw} (tacitly assumed to be a function of x) to express the coefficient with which the local velocity has to be multiplied to find the associated (shifted) pressure, in an analogous way as Eq. (3). The function X_{ffw} is recognized to be a transfer impedance. In essence, it is a continuous function of x but as before we will consider it as a finite sequence of N samples. In matrix language X_{ffw} is to be converted into an *off-diagonal* matrix $\mathbf{X}_{\text{ffw}}$; the diagonal is distant from the main diagonal by a number of steps that corresponds to Δx. When $\mathbf{X}_{\text{ffw}}$ contains elements below the main diagonal, we have feed-forward, when it contains elements above the main diagonal, we have feed-backward. If Eq. (3) is rewritten as $p=-\frac{1}{2}Z_{\text{eff}}v$, it is easy to see that the product $\frac{1}{2}\mathbf{X}_{\text{ffw}}\mathbf{v}$ relates the (shifted) OHC-generated pressure to the velocity $\mathbf{v}$. The model equation then becomes

$$\left(i\omega\rho\mathbf{G} + \tfrac{1}{2}\mathbf{X}_{\text{ffw}}\right)\mathbf{v} = -i\omega\rho\mathbf{S}v_{\text{st}}, \tag{4}$$

in complete analogy to Eq. (1).[2] In the forward solution, it is this equation that has to be solved for $\mathbf{v}$.

In the feed-forward case, the first step of the inverse solution is identical to that in a classical model: the pressure $\mathbf{p}$ is computed from the hydrodynamic equation (2). Note again that this step only involves hydrodynamical concepts, the matrix $\mathbf{G}$ and the vector $\mathbf{S}$. The second step of the inverse solution is different because the vectors $\mathbf{v}$ and $\mathbf{p}$ do not refer to the same points on the x-axis. Dividing the *shifted* pressure p by the unmodified velocity v produces the function X_{ffw}, in an analogous fashion as Z_{eff} in Eq. (3). This is the feed-forward impedance function which, transformed into an off-diagonal matrix $\mathbf{X}_{\text{ffw}}$ and used in the model Eq. (4), should produce an accurate replica of the original given BM velocity $\mathbf{v}$. This elaboration forms the core of the feed-forward inverse solution and the associated resynthesis.

Actually, the situation is somewhat more complicated. We must take into account that the BM in the model has an inherent passive mechanical component that is not influenced by actions of OHCs. For reasons of simplicity, we will assume that this component is not involved in any form of feed-forward and that it can be represented by a classical impedance, called the passive BM impedance. This impedance has to be included in our inverse procedure but it has to remain outside the feed-forward process. Call that impedance function Z^{pass} and its associated diagonal matrix $\mathbf{Z}^{\text{pass}}$. Redefine the pressure $\mathbf{p}$ as follows:

$$\mathbf{p} = \left(i\omega\rho\mathbf{G} + \tfrac{1}{2}\mathbf{Z}^{\text{pass}}\right)\mathbf{v} + i\omega\rho\mathbf{S}v_{\text{st}}. \tag{5}$$

It is this pressure that has to be produced by the feed-forward impedance X_{ffw}. Equation (5) is the analog of Eq. (2) for the case where the classical passive BM impedance is explicitly included. The pressure $\mathbf{p}$ according to Eq. (5) is, in shifted form, divided by $\mathbf{v}$ to yield the transfer impedance X_{ffw}. Convert this impedance to an off-diagonal matrix $\mathbf{X}_{\text{ffw}}$ and substitute it into the "complete" model equation:

$$\left(i\omega\rho\mathbf{G} + \tfrac{1}{2}(\mathbf{Z}^{\text{pass}} + \mathbf{X}_{\text{ffw}})\right)\mathbf{v} = -i\omega\rho\mathbf{S}v_{\text{st}}. \tag{6}$$

This equation can be solved for $\mathbf{v}$. Note that $\mathbf{Z}^{\text{pass}}$ is a diagonal and $\mathbf{X}_{\text{ffw}}$ an off-diagonal matrix. The so-obtained resynthesized response should be a good approximation of the given BM velocity with which the inverse procedure was started. It has been our experience that in the feed-forward case the inverse method is more prone to errors than in the classical case. Therefore, the resynthesis step is really essential.

From the above derivation, it should be clear that $\mathbf{Z}^{\text{pass}}$ does not necessarily have to be the passive BM impedance, measured *postmortem*. It can be any impedance that behaves as a point-impedance, i.e., any impedance that relates pressure to velocity at the same location. In fact, it can be a combination of the actual passive BM impedance and a fraction of the (classical) active component of the effective BM impedance. In that way any "degree" of feed-forward can be built in (from 0 to 100%), and in that respect, the inverse-solution method outlined above is quite universal. In the present paper, we will leave that extension unexplored, to preserve space, and we will consider only the passive component of the BM impedance to be separated out in the classical sense as in Eqs. (5) and (6).

III. EXAMPLE I: FEED-FORWARD—SINGLE TONES

Responses were measured at the BM of the guinea pig, with the beam of a laser interferometer directed at a location on the BM that has a best frequency (BF) in the range from 16 to 20 kHz. The measurement method has been described before (de Boer and Nuttall, 2000). The protocols of the experiments, which were performed on deeply anesthetized animals, were approved by the Oregon Health & Science Committee on the Use and Care of Animals and consistent with NIH guidelines for humane treatment of animals. For the single-tone stimulus case, we utilized responses measured at low levels, of the order of 20–40 dB sound pressure level (SPL), and responses of the same cochlea measured *postmortem*. Responses are measured as functions of frequency and have been converted into "cochlear patterns,"

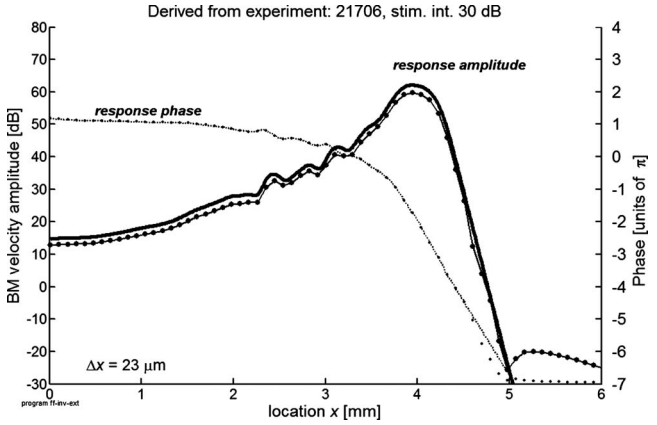


FIG. 1. Result of resynthesis in the single-tone case. Length of the model: 12 mm, divided into 1024 segments. Parameters of model are derived from an experiment. Frequency: 17 kHz. The passive component of the BM impedance has been determined from the *postmortem* response of the same animal, and the remainder of BM dynamics is interpreted in terms of feed-forward (see Eqs. (5) and (6)). Full-drawn and finely dotted curves: original response, amplitude (scale on the left), and phase (scale on the right). Coarsely dotted curves: resynthesized response, amplitude, and phase. For clarity, the amplitude of the resynthesized response has been plotted 2 dB lower than the original amplitude. This figure illustrates the fidelity of the sequence inverse-solution-resynthesis in the feed-forward model.

i.e., response functions as functions of the location variable x, valid for one frequency (usually chosen to be equal to the BF corresponding to a low-level response). A standard frequency-location map has been used in the conversion, refinements of the mapping procedure are described in de Boer and Nuttall (1999). For the two-tone computations (Sec. IV), we assume that the model is stimulated by two fairly strong tones of which we know the excitation patterns. In that case, we estimate the distribution of DP sources for third-order distortion occurring in the OHCs—in the same way as we have done before (de Boer *et al.*, 2007).

By way of Fig. 1, we illustrate the fidelity of the inverse solution in the case where one component of BM dynamics is contained in the *postmortem* impedance Z^{pass} and the other part involved in the feed-forward process. That is, we have used Eq. (5) as the basis of the inverse solution and Eq. (6) for resynthesis. The model has a length of 12 mm (of which only a part is shown), divided into 1024 sections. The feed-forward distance Δx has been given the value 23 μm. For clarity the resynthesized amplitude is drawn 2 dB below the original amplitude curve; the two phase curves coincide over most of their ranges, and are seen to depart from one another where the amplitude has dropped more than 50 dB from its peak.

The example of Fig. 1 has been selected because of its almost ideal resynthesis. More realistic cases are presented by Fig. 2 which shows results of four experiments, in four different animals. Again it is evident that the resynthesized response curves are good imitations of the original responses. However, there is more variability in and between the data sets, but that is mainly outside the main response region. We conclude that the inverse solution is sufficiently accurate for our further needs. We will now use this method to predict the behavior of DP waves in a feed-forward setting.

IV. EXAMPLE II: FEED-FORWARD—TWO STIMULUS TONES AND DPS

Let us consider the case where the model is stimulated by two tones, with frequencies f_1 and f_2. These tones are presented with an intermediate intensity; we will call the appropriate response functions $v^{(1)}$ and $v^{(2)}$. Again we have

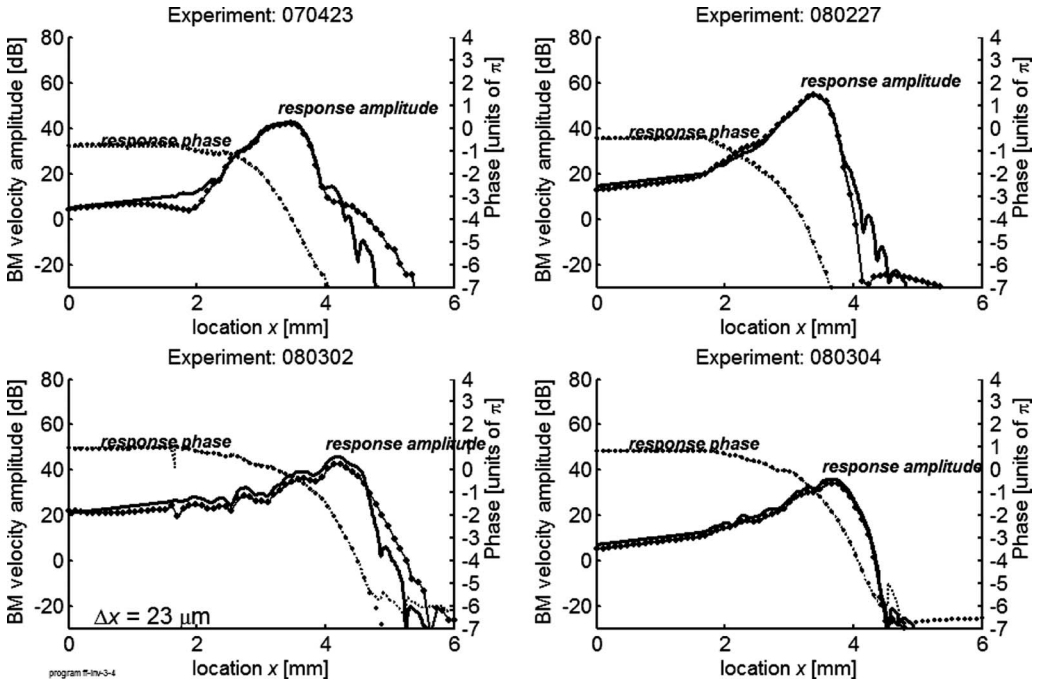


FIG. 2. Results of inverse solution followed by resynthesis in four different experiments. Parameters of the model are derived from these experiments. Tone frequency: 17 kHz. The amplitude of the resynthesized wave is drawn 2 dB lower than the original amplitude.

omitted the dependence on location x. For each of the two tones, the corresponding transfer impedance X_{ffw} can be determined; call these two functions $X_{\mathrm{ffw}}^{(1)}$ and $X_{\mathrm{ffw}}^{(2)}$. It is repeated that both functions only represent the OHC-controlled components of the BM dynamics; the passive component Z^{pass} has been separated out. The transfer impedances $X_{\mathrm{ffw}}^{(1)}$ and $X_{\mathrm{ffw}}^{(2)}$ as well as the passive components Z^{pass} refer to the two different frequencies, f_1 and f_2, respectively. A product $p_{\mathrm{act}}^{(1)}$ like

$$p_{\mathrm{act}}^{(1)} = -\tfrac{1}{2}\mathrm{v}^{(1)}X_{\mathrm{ffw}}^{(1)} \tag{7}$$

will now represent the active pressure developed by the hair cells due to tone 1 for the frequency f_1. Similarly, $p_{\mathrm{act}}^{(2)}$ defined by

$$p_{\mathrm{act}}^{(2)} = -\tfrac{1}{2}\mathrm{v}^{(2)}X_{\mathrm{ffw}}^{(2)} \tag{8}$$

will represent the OHC pressure corresponding to tone 2 for the frequency f_2. In both products, the displacement Δx has to be taken into account, of course.[3] In the limit of small nonlinear distortion, the DP with frequency $2f_1-f_2$ will appear in the output of the OHCs with an amplitude $A_{\mathrm{DP}}(x)$ given by (for more details, see de Boer *et al.*, 2007)

$$A_{\mathrm{DP}}(x) = A(p_{\mathrm{act}}^{(1)}(x))^2 \,.\, (p_{\mathrm{act}}^{(2)}(x))^*, \tag{9}$$

where we have deliberately reintroduced the independent variable x. It is recalled that $p_{\mathrm{act}}^{(1)}(x)$ and $p_{\mathrm{act}}^{(2)}(x)$ refer to the same x-axis but intrinsically contain information about two different frequencies, f_1 and f_2. In contrast, the resulting coefficient $A_{\mathrm{DP}}(x)$ describes the complex amplitude for a wave with the DP frequency f_{DP} equal to $2f_1-f_2$. The square in Eq. (9) arises because the expression for f_{DP} contains the term $2f_1$, the superscript asterisk (*) denotes the complex conjugate and is needed because the DP frequency contains the term $-f_2$. The multiplier A will be specified further on. It is stressed that Eq. (9) implies that the drive behind the DP wave, as it is expressed by the pressure $A_{\mathrm{DP}}(x)$, is solely due to physiological effects produced by viable OHCs (as in our earlier work on DP waves, de Boer *et al.*, 2007). That we are again using the case of small nonlinear distortion is based on the fact that experiments on DP waves have yielded consistent results for stimulus signals as weak as 40 and 50 dB SPL (He *et al.*, 2007).

The pressure distribution $A_{\mathrm{DP}}(x)$ acts as a source distribution for the resulting DP wave, with the DP frequency equal to $2f_1-f_2$. That DP wave can now be computed from a small variation in the model equation, Eq. (6), where the stapes excitation term $-i\omega\rho\mathbf{S}v_{\mathrm{st}}$ is replaced by the column vector $\mathbf{A}_{\mathrm{DP}}$ which represents $A_{\mathrm{DP}}(x)$. The resulting equation reads

$$\left(i\omega\rho\mathbf{G} + \tfrac{1}{2}(\mathbf{Z}^{\mathrm{pass}} + \mathbf{X}_{\mathrm{ffw}})\right)\mathbf{v} = +\mathbf{A}_{\mathrm{DP}}. \tag{10}$$

All terms are to be evaluated at the DP frequency f_{DP}. The transfer impedance X_{ffw} is determined from a measured tone response of the animal, at the lowest feasible level (usually 20 dB SPL). The passive impedance Z^{pass} is determined from the *postmortem* response, at the highest level (100 dB). Both of these parameters are evaluated for the DP frequency and are functions of location x. For the computation of $A_{\mathrm{DP}}(x)$ (see Eq. (9)), responses and impedances are used for higher levels, adapted to the appropriate frequencies, f_1 and f_2. For the method to reduce reflection of the DP wave at the stapes end, we refer to de Boer *et al.* (2007).

Figure 3 shows four representative theoretical results computed from results of four experiments. Two tones of 40 dB SPL with frequencies f_1 and f_2, with the frequency ratio equal to 1.08, were used as stimuli. The amplitudes of the excitation patterns produced by these tones are shown by dashed curves (labeled: f_1 and f_2) and suggested by shading. The levels at which these patterns are plotted are arbitrary; they are selected for visual clarity. The resulting DP waves are computed as described above. In each panel, the solid thick curve shows the amplitude of the computed DP wave and the coarsely dashed curve shows the phase; the scale for the amplitude appears on the left, that for the phase on the right. The amplitude factor A in Eq. (9) has been selected so that in the region of maximum response to the DP frequency, the amplitude is around 50 dB.

Figure 3 shows results for the same experiments as used for the publication at the Keele Symposium (de Boer and Nuttall, 2008), produced with a small variation in the original program. Figure 4 shows results, plotted in the same layout, for a further set of four animals. In this case the results show somewhat more variability. It should be remembered that the "best" results of computations are usually obtained from experiments that yield "smooth" impedance curves; in the present cases more realistic results are shown. It should be clear that a considerable part of the DP wave propagates in the forward direction. In addition, this property is evident also outside the region where the amplitude of the DP wave is substantial, and that is the main result of the present study. There are irregularities in the phase curves but in no case is a clear reverse wave present near or not far from the stapes. As an afterthought, it is quite surprising that a feed-forward effect, active over a distance of just a few hair-cell diameters, and only present in the active region of the cochlea, can have such a profound effect, even on small-amplitude waves at a considerable distance.

V. ABOUT THE DP COMPRESSION WAVE

We were inspired to study feed-forward (and its variations) by the results from measurements on the DP wave, both inside the cochlea (see the references mentioned in Sec. I) and outside it, in the form of otoacoustic emissions (Ruggero, 2004; see also Siegel *et al.*, 2005). As stated in Sec. I, it has been attempted by some authors to explain these findings by a mechanism that invokes a compression wave. That explanation has been criticized on several grounds, notably on its inability to explain frequency-and-location-selective "activity" (de Boer and Nuttall, 2006a) and its failure to explain the Allen–Fahey effect (Shera *et al.*, 2007; Shera and Guinan, 2007). In other words, it cannot be the compression wave that induces propagation of DP waves in the cochlea. (To quote Shakespeare's Hamlet: "There are more things in heaven and earth, Horatio, than are dreamt of in your philosophy.") In all that work, the background of the elabora-

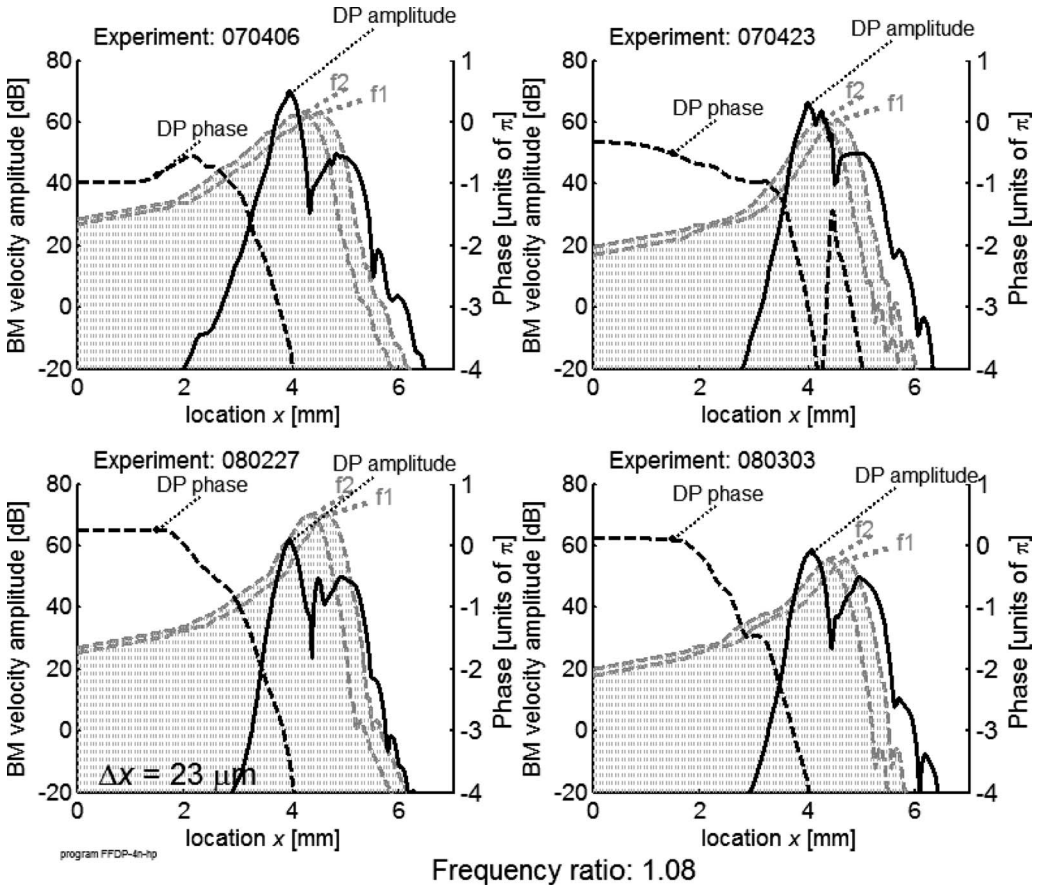


FIG. 3. DP waves generated in a stylized three-dimensional model with feed-forward. Two tones of 40 dB SPL with frequencies f_1 (14.57 kHz) and f_2 (15.74 kHz) are used as stimuli. The amplitudes of the excitation patterns produced by these tones are shown by dashed curves (labeled f_1 and f_2) and suggested by shading. Generation of the DP wave with frequency $2f_1-f_2$ (13.41 kHz) computed as described in de Boer *et al.* (2007), and further described in the text. Reflection of the DP wave at the stapes has been suppressed. Solid thick curve: amplitude of the resulting DP wave. Ordinate scale on the left. Coarsely dashed curve: the associated DP phase response, scale on the right. The thick and the thin dotted curves show amplitude and phase, respectively, of the response to a single low-level tone with the DP frequency started at the stapes. These curves serve as landmarks.

tions has been a classical model of the cochlea. We presently want to discuss the effect of a compression wave in a feed-forward model, but we should keep in mind the reservations expressed in the cited papers.

The main point is this: In the cases we have shown in Figs. 3 and 4 the computed DP wave generated by the feed-forward mechanism is extremely small in the region between the stapes and the primary-tone response patterns. It is so small that whatever is the amplitude of the *compression-induced wave* (in the sequel to be called CI-W), the latter will be dominant in most of that region. This type of DP wave would then be the only source of otoacoustic emissions and also control what we observe inside the cochlea. For this reason (and for this reason only, see the above-mentioned counter-arguments and restrictions), we will present an estimate of the CI-W in our model.

The compression-induced DP wave has been analyzed in Shera *et al.* (2007), and we accept the main idea. We will assume that $A_{\mathrm{DP}}(x)$ given by our Eq. (9) is also the source distribution of the compression wave. In this way, we will be able to compare the CI-W directly with the DP wave considered in Sec. IV.

Two compression waves arise from the overlap region, one toward the helicotrema and one toward the stapes. In theoretical exercises, the fluid is usually considered incompressible. In a more practical situation, the fluid is compressible and the speed of sound in it finite. In fact, for frequencies around 17 kHz, the length of the guinea-pig cochlea is of the order of a quarter wavelength. This means that for a high frequency the basally generated compression wave encounters an obstacle. The essentially very large impedance of the top of the cochlea (it is all bone) is transformed into a small impedance in the region where the compression wave is generated. Therefore, we must expect that the apical part of the compression wave "shunts" the basal part so that the latter arises from a smaller pressure than $A_{\mathrm{DP}}(x)$. If we take $A_{\mathrm{DP}}(x)$ as the source and compute a compression-induced DP wave from it, we are decidedly *over-estimating* the basal compression wave. Here we end our short excursion into the domain of finite compressibility of the fluid. In the sequel, we consider only an ideal fluid.

Integration over space yields the pressure p_{evo} that drives the compression wave:

$$p_{\mathrm{evo}} = \int_0^L A_{\mathrm{DP}}(x)dx, \tag{11}$$

where L denotes the length of the model.[4] As said, from here

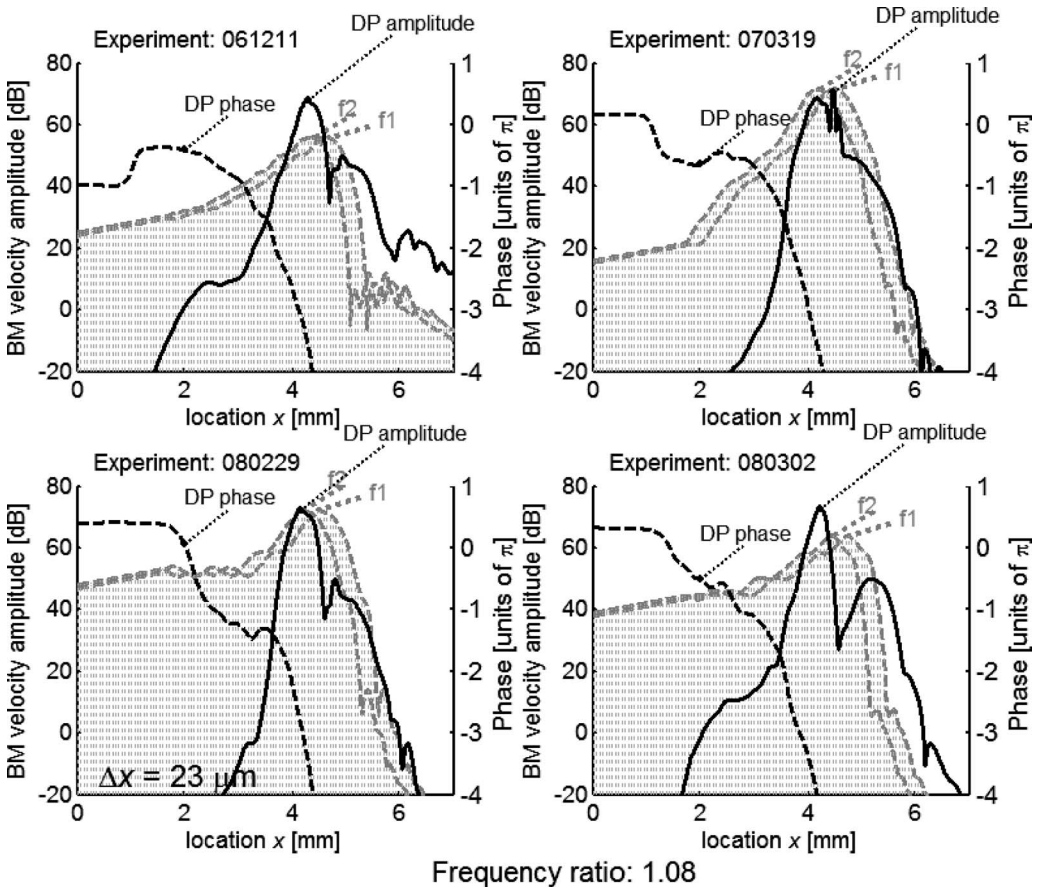


FIG. 4. Same as Fig. 3, with a different selection of animals. In these cases, the results display more variability.

on we again assume the compression wave to have an infinite velocity of propagation; hence p_{evo} is the same everywhere. At the location of the stapes, the compression wave is equal to p_{evo}. If we divide the pressure p_{evo} by the effective input impedance of the cochlea (the impedance loading the oval window), we obtain the equivalent stapes velocity (at least, if we ignore the impedance of the round window, see Shera *et al.*, 2007). The input impedance of the model can easily be found from our computations. To this impedance, another impedance must be added: the radiation impedance representing the sound that is going to the outer world, via the middle ear. For simplicity, we assume this impedance to be equal to the aforementioned input impedance—implying an ideal impedance match between cochlea and middle ear. This reduces the amplitude of the CI-W by 6 dB. It should finally be recalled that all variables involved must be computed for the DP frequency.

Figure 5 shows four representative results. The figure shows the amplitudes of the computed CI-Ws. These amplitudes are normalized in exactly the same way as the DP waves. As a result, the panels display the relation between the CI-W and the "regular" feed-forward DP wave. It is clear that there is considerable variation between the panels; we have selected experiments from the sets shown in earlier figures to demonstrate the extremes of the CI-W. It is also clear that in the region where the DP frequency shows its own maximum, the DP wave and the CI-W are of comparable magnitude. The figure confirms what has been said above: outside the main region of the DP wave it is the CI-W that dominates. With respect to direction of propagation, there is no difference. However, because the CI-W does not have the required properties (see above), we should see these results only as indicative, not as final outcomes. At any rate, the figures show the relation between the amplitudes of the computed CI-W and the regular feed-forward DP wave correctly, because they have been derived from the same DP source function $A_{DP}(x)$ and plotted with the same normalization factor. We should recall that the compression wave can directly be identified in experiments on intra-cochlear pressure (Dong and Olson, 2008 and earlier work). It was found that the Distribution Product Oto-Acoustical-Emission (DPOAE) related to the compression wave is "Much lower than the actual DPOAE we have measured in the ear canal" (citation from *loc. cit.*).

VI. EXPLANATIONS, CONCLUSIONS, LIMITATIONS

The main property that emerges from Figs. 3 and 4 is that the computed DP wave is almost exclusively a forward wave, over most of its range. The theory described in de Boer (2007) yields that forward and reverse waves have quite different amplification properties, but that theory is only valid in a structure exclusively governed by feed-forward. Apparently, it is also valid in the case where the passive component of the BM impedance is included.

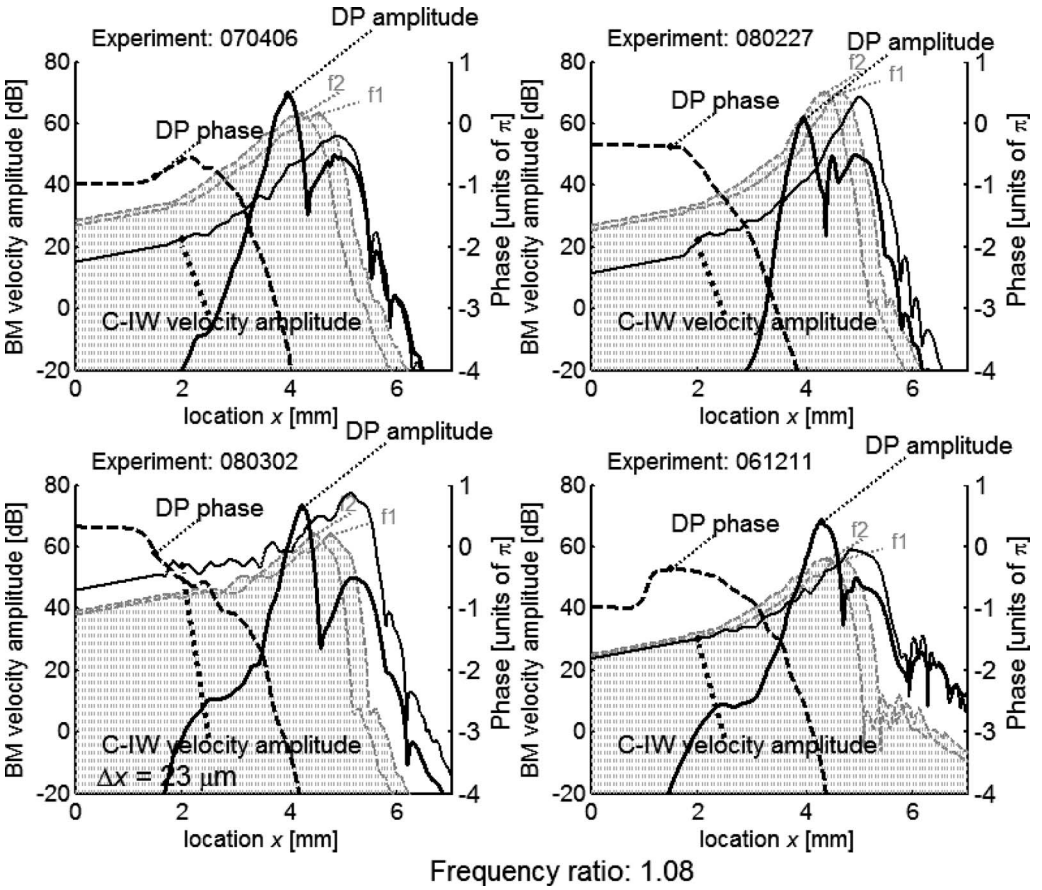


FIG. 5. In addition to the DP wave, the CI-W is shown. Only the amplitude of that wave is shown (the phase follows that of a simple tone at the same frequency injected at the stapes). Experiments have been selected to show "normal" as well as "large" amplitudes of that wave. It has been assumed that the middle ear is loaded by the same impedance as the input impedance of the cochlea. The short-circuit effect of the apical part of the CI-W (see text) has been neglected.

Where the DP wave originates, in the region of overlap, both forward and reverse waves are initiated. Going to the left, toward the stapes, the reverse wave starts to be progressively attenuated in a feed-forward model. And still more to the left even the forward wave dwindles to a very small amplitude. One may ask the following questions: Why is not there a substantial forward wave in this region? Does not there exist some kind of effective stimulus here that would generate it? The answer is no. In this region, the DP drive pressure $A_{DP}(x)$ is too small to have an effect by itself. Secondary effects (due to hydrodynamics and the passive BM impedance) of that drive pressure have been found to remain restricted to the region where $A_{DP}(x)$ is maximal.

In effect, the theoretical prediction of a small forward-going DP wave in this region is contrary to experiment. In de Boer *et al.* (2008), it is shown that the forward-propagating character of the DP wave is present in almost all of the region between overlap region and stapes. A similar conclusion can be drawn from the other papers on this subject mentioned in Sec. I. Apparently, if feed-forward is a concept to be included in the explanation of the data, it has to be extended or modified. We recall what has been said about a possible combination of a classical and a feed-forward model at the end of Sec. II.

There is a further problem with the feed-forward concept. As demonstrated here, the principle of feed-forward does not allow a sizable reverse DP wave to occur. The same would be true for any other (non-classical) type of reverse wave. As a consequence, the principle of coherent reflection cannot work in such a model. We further refer to the discussion of this topic in the Keele paper (de Boer and Nuttall, 2008), and await further theoretical developments. At this point, we can only add that a mechanism including a CI-W does not alleviate that problem. Only the extension of the feed-forward concept alluded to in the ending part of Sec. II can do that.

For the sake of completeness, we computed the CI-W, and found it to have all the properties we could have predicted. We did not pursue this topic any further because, from various sides, it can be argued that the CI-W cannot be the DP wave observed inside the cochlea and cannot be the source of DP otoacoustic emissions.

ACKNOWLEDGMENTS

From the many data sets collected in close collaboration with Jiefu Zheng, a number of the smoothest responses were selected for the illustrations in this paper. Dr. Zheng is gratefully acknowledged for his important contributions. Interactions with Christopher Shera have considerably improved the authors' thinking. The authors are also grateful to Renato

Nobili and Fabio Mammano who kindly put their Green's function program code at their disposal. This study received support from NIH NIDCD Contract No. R01 DC 00141.

[1]The minus sign and the factor 2 are due to sign conventions. Upward movement of the BM is counted positive, p is the pressure in the upper channel, and $-p$ in the lower channel.

[2]Actually, in the resynthesis procedure, the stapes velocity should be replaced by the virtual stapes velocity (de Boer and Nuttall, 1999).

[3]It should be noted that the expression $\frac{1}{2}v^{(1)}X^{(1)}_{\mathrm{ffw}}$ with shifted x-scales in the two factors is identical to the corresponding classical expression $\frac{1}{2}v^{(1)}Z^{(1)}_{\mathrm{BMact}}$ with equal x-scales, where $Z^{(1)}_{\mathrm{BMact}}$ is the "active" component of the classical BM impedance for tone 1. This can be seen by comparing Eq. (6) with its classical counterpart, Eq. (1). The same is valid for tone 2.

[4]At first sight, it may be confusing to observe that $A_{\mathrm{DP}}(x)$ and p_{evo} have different dimensions. In fact, $A_{\mathrm{DP}}(x)$ has the meaning of being a pressure *density*, the pressure produced by OHCs per unit of length.

Allen, J. B. (**1980**). "Cochlear micromechanics—A physical model of transduction," J. Acoust. Soc. Am. **68**, 1660–1670.

de Boer, E. (**1993**). "The sulcus connection. On a mode of participation of outer hair cells in cochlear mechanics," J. Acoust. Soc. Am. **93**, 2845–2859.

de Boer, E. (**1995a**). "The "inverse problem" solved for a three-dimensional model of the cochlea. I. Analysis," J. Acoust. Soc. Am. **98**, 896–903.

de Boer, E. (**1995b**). "The "inverse problem" solved for a three-dimensional model of the cochlea. II. Application to experimental data sets," J. Acoust. Soc. Am. **98**, 904–910.

de Boer, E. (**1996**). "Mechanics of the cochlea: Modeling efforts," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 258–317.

de Boer, E. (**2007**). "Forward and reverse waves in nonclassical models of the cochlea," J. Acoust. Soc. Am. **121**, 2819–2821.

de Boer, E., and Nuttall, A. L. (**1999**). "The "inverse problem" solved for a three-dimensional model of the cochlea. III. Brushing up the solution method," J. Acoust. Soc. Am. **105**, 3410–3420.

de Boer, E., and Nuttall, A. L. (**2000**). "The mechanical waveform of the basilar membrane. III. Intensity effects," J. Acoust. Soc. Am. **107**, 1497–1507.

de Boer, E., and Nuttall, A. L. (**2002**). "The mechanical waveform of the basilar membrane. IV. Tone and noise stimuli," J. Acoust. Soc. Am. **111**, 979–989.

de Boer, E., and Nuttall, A. L. (**2006a**). "Amplification via "compression waves" in the cochlea—A parable," *Abstracts MidWinter Meeting*, Association for Research in Otolaryngology, Baltimore, MD, USA, Vol. **29**, p. 349.

de Boer, E., and Nuttall, A. L. (**2006b**). "Spontaneous basilar-membrane oscillation (SBMO) and coherent reflection," J. Assoc. Res. Otolaryngol. **7**, 26–37.

de Boer, E., and Nuttall, A. L. (**2008**). "Obvious and 'hidden' waves in the cochlea," in *Mechanics of Hearing*, edited by N. P. Cooper and D. T. Kemp (World Scientific, Singapore), pp. 34–39.

de Boer, E., Nuttall, A. L., and Shera, C. A. (**2007**). "Wave propagation patterns in a "classical" three-dimensional model of the cochlea," J. Acoust. Soc. Am. **121**, 352–362.

de Boer, E., Zheng, J., Porsov, E., and Nuttall, A. L. (**2008**). "Inverted direction of wave propagation (IDWP) in the cochlea," J. Acoust. Soc. Am. **123**, 1513–1521.

Dong, W., and Olson, E. S. (**2008**). "Supporting evidence for reverse cochlear traveling waves," J. Acoust. Soc. Am. **123**, 222–240.

Geisler, C. D., and Sang, C. (**1995**). "A cochlear model using feed-forward outer-hair-cell forces," Hear. Res. **86**, 132–146.

He, W., Fridberger, A., Porsov, E., Grosh, K., and Ren, T. (**2008**). "Reverse wave propagation in the cochlea," Proc. Natl. Acad. Sci. U.S.A. **105**, 2729–2733.

He, W., Nuttall, A. L., and Ren, T. (**2007**). "Two-tone distortion at different longitudinal locations on the basilar membrane," Hear. Res. **228**, 112–122.

Mammano, F., and Nobili, R. (**1993**). "Biophysics of the cochlea: Linear approximation," J. Acoust. Soc. Am. **93**, 3320–3332.

Neely, S. T., and Kim, D. O. (**1986**). "A model for active elements in cochlear biomechanics," J. Acoust. Soc. Am. **79**, 1472–1480.

Patuzzi, R. (**1996**). "Cochlear micromechanics and macromechanics," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 186–257.

Ren, T. (**2004**). "Reverse propagation of sound in the gerbil cochlea," Nat. Neurosci. **7**, 333–334.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian ear," Physiol. Rev. **81**, 1305–1352.

Ruggero, M. A. (**2004**). "Comparison of group delay of $2f_1-f_2$ distortion product otoacoustic emissions and cochlear travel times," ARLO **5**, 143–147.

Shera, C. A., and Guinan, J. J. (**2007**). "Cochlear traveling-wave amplification, suppression, and beamforming probed using noninvasive calibration of intracochlear distortion sources," J. Acoust. Soc. Am. **121**, 1003–1015.

Shera, C. A., Tubis, A., Talmadge, C. L., de Boer, E., Fahey, P. A., and Guinan, J. J. (**2007**). "Allen–Fahey and related experiments support the predominance of cochlear slow-wave otoacoustic emissions," J. Acoust. Soc. Am. **121**, 1564–1575.

Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (**2005**). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," J. Acoust. Soc. Am. **118**, 2434–2443.

Steele, C. R., Baker, G., Tolomeo, J., and Zetes, D. (**1993**). "Electromechanical models of the outer hair cell," in *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 207–214.

Steele, C. R., and Lim, K.-M. (**1999**). "Cochlear model with three-dimensional fluid, inner sulcus and feed-forward mechanism," Audiol. Neuro-Otol. **4**, 197–203.

Wen, B., and Boahen, K. (**2003**). "A linear cochlear model with active bi-directional coupling," in Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Cancun, Mexico, Vol. **3**, pp. 2013–2016.

Zwislocki, J. J., and Kletsky, E. J. (**1979**). "Tectorial membrane: A possible effect on frequency analysis in the cochlea," Science **204**, 639–641.

Judging the similarity of soundscapes does not require categorization: Evidence from spliced stimuli

Jean-Julien Aucouturier[a)]
College of Liberal Arts, Temple University, Japan Campus, 2-8-12 Minami Azabu, Minato-ku, 106-0047 Tokyo, Japan

Boris Defreville[b)]
ORELIA, 17 Rue de la Cloche, 77300 Fontainebleau, France

(Received 27 February 2008; revised 20 January 2009; accepted 26 January 2009)

This study uses an audio signal transformation, *splicing*, to create an experimental situation where human listeners judge the similarity of audio signals, which they cannot easily categorize. Splicing works by segmenting audio signals into 50-ms frames, then shuffling and concatenating these frames back in random order. Splicing a signal masks the identification of the categories that it normally elicits: For instance, human participants cannot easily identify the sound of cars in a spliced recording of a city street. This study compares human performance on both normal and spliced recordings of soundscapes and music. Splicing is found to degrade human similarity performance significantly less for soundscapes than for music: When two spliced soundscapes are judged similar to one another, the original recordings also tend to sound similar. This establishes that humans are capable of reconstructing consistent similarity relations between soundscapes without relying much on the identification of the natural categories associated with such signals, such as their constituent sound sources. This finding contradicts previous literature and points to new ways to conceptualize the different ways in which humans perceive soundscapes and music.
 [DOI: 10.1121/1.3083232]

PACS number(s): 43.66.Ba, 43.50.Rq, 43.75.Cd, 43.60.Cg [DD] Pages: 2155–2161

I. INTRODUCTION

Categorizing and comparing percepts are central abilities for human cognition. They do not operate in isolation, but rather depend greatly on one another. When we lack specific knowledge, we often rely on domain-general similarity to guide categorization.[1] Reciprocally, we often judge the similarity of objects by comparing their features, i.e., based on their prior categorization.[2] However, the exact nature of the relation between similarity and categorization still eludes modern cognitive science.[3] There have been empirical demonstrations of categorization processes conducted in apparent independence from similarity, but rather based on semantic representations looking like rules or theories.[4] Further, features used for similarity depend on context, culture, and experience, and different strategies often operate at the same time to reinforce each other and eliminate context-driven noise.

This complex interdependence also holds in the specific case of auditory perception. The vast psychology literature studying similarity judgments for musical signals has made constant efforts to control for unwanted high-level recognition effects, by using short or synthesized stimuli.[5,6] Research on the perception of environmental audio scenes (or *soundscapes*) also suggests that similarity and categorization are deeply entangled. The perceived similarity of soundscapes is believed to be based on the prior identification of their constituent sources:[7,8] "If there are cars in both signals, then these are similar." Similarly, psycholinguistic studies[9] have concluded that the subjective "unpleasantness" of an urban soundscape increases when more mechanical sound sources (e.g., vehicles) are identified than natural sources (e.g., voices or birds).

In the case of soundscapes, however, recent computational studies seem to contradict that similarity should rely on categorization. A common algorithmic paradigm to simulate human judgments of audio similarity is to model audio signals as global statistical distributions of local features computed on very short (cf. 50-ms) frames. This technique, called bag-of-frames (BOF), uses only low-level and short-term spectral information, and does not rely on any prior categorization of, e.g., sound sources. In a recent study,[10] we established that BOF, simple as it is, was computationally sufficient to simulate similarity with near-perfect precision for soundscapes. Therefore, from the sole point of view of information processing, categorization does not seem to be necessary to compare soundscape signals. However, we found significantly lower performance when applied to polyphonic music signals.

The present paper aims to clarify this situation. While our previous study[10] was a machine experiment, this one is a human psychological experiment. We create an experimental situation where human listeners produce judgments of similarity for audio signals, which they cannot easily categorize, by manipulating the signals with an algorithmic transformation called *splicing*. Splicing alters an audio signal by segmenting it into 50-ms frames, then shuffling and concatenat-

a)Electronic mail: aucouturier@gmail.com
b)Electronic mail: boris.defreville@orelia.fr

ing these frames back in random order. Spliced signals by construction have the same frames than the original signals, hence the same local features, as well as the same statistics of features; only the time ordering of the frames is modified.

Splicing is a known technique in language science: It is typically used to process spoken sentences for listening experiments so that their lexical content is lost but their vocal content (prosody, timbre, etc.) preserved to some extent.[11] Similar effects are known for music signals: It is difficult to identify instruments or even produce any qualitative description of shuffled or reversed music.[12] This is hardly surprising: Psychophysical experiments[5] have long established the importance of dynamics, notably the attack time and fluctuations of the spectral envelope, in the perception of individual instrument notes.

In this study, we confirm that splicing acts as a categorization mask for both music and soundscape, and use this property to study the influence of categorization on the processing of similarity for both music and soundscapes.

II. METHODS

A. Data

Each subject was presented (in tasks described below) a set of 12 audio stimuli in either of 4 conditions: normal music (NM), spliced music (SM), normal soundscape (NS), spliced soundscape (SS). Each stimulus is a 20-s stereo sound file sampled at 44.1 kHz, normalized at zero dc, and has maximum amplitude of −3 dB. The 12 stimuli in the NM condition are excerpts of western popular music of various genres from the period 1950–2000. The 12 stimuli in the NS condition are excerpts of urban soundscapes used in our previous study.[10] Stimuli in the two spliced conditions (SM and SS) were obtained from the stimuli in the associated normal condition (NM and NS, respectively) by the following algorithmic transform:

(1) A "normal" signal is truncated into successive, nonoverlapping 2048-point time frames (46.4 ms at 44.1 kHz).
(2) Frames are shuffled in random order (using a pseudorandom generator initialized with a common seed for all files).
(3) Frames are concatenated back to an audio signal of same length as original.

B. Participants

55 subjects participated, all relatively young ($M=20.70$, $SD=3.56$) Japanese undergraduate students. There were 45 male participants, and 10 female participants, with no significant age difference between sexes [$F(1,55)=1.15$, $p=0.288$]. All participants distributed in roughly equal proportion over all four conditions (NM:15, SM:15, NS:13, and SS:12), with no gender bias. We kept data from all participants for the analysis presented below.

C. Tasks

Each subject was presented stimuli from only one of the four tested conditions (randomized). The test consisted of a categorization task (used to validate the effect of splicing) and a similarity task. To prevent priming effects, every subject performed the similarity task before the categorization task.

1. Similarity task

In the similarity task, participants were asked to judge similarity relations in a set of triads presented one after the other. We presented 44 triads per condition, arranged in a $\lambda=2$ balanced incomplete block design (all possible pairs of stimuli appear twice). Although the set of triads is incomplete, it was designed so that all triads in a given spliced condition have a corresponding normal triad in the normal condition composed of the three same stimuli in their original version.

For each presented triad, participants were asked to select the pair of stimuli they deemed were most similar to one another. Participants were given no other information than the sole audio presentation of the stimuli (no file name, no editorial information on the recording, etc.). Prior to the test, participants received the following instructions:

> When judging the "similarity" of such sounds, please try to listen to the immediate, global impression of the sound, and not to finer, analytical details.[13] For instance, you may want to say that a Bob Dylan song "sounds like" another piece (e.g., a Bossa Nova piece by Joao Gilberto), because they both have a soft acoustic guitar and a gentle male voice. In a noisy background, e.g., a crowded cafe, one could be mistaken for the other. Judging that the same Bob Dylan guitar song is similar to, say, a Joni Mitchell piano song (because both artists have the same influences, and performed at similar periods), or that the Bossa Nova piece is similar to a recent Brazilian electronic piece (because they are both from Brazil) would be irrelevant for this test.

Additionally, when tested in one of the spliced conditions (SM and SS), participants were instructed to do the following:

> [...] please try to listen to these sound extracts as if they were natural ones, i.e., focus on your immediate, holistic sensation of how much these sounds sound like one another (and try to refrain from analytical thinking).

Stimuli could be listened to as many times as needed; answers could be given before listening to a stimuli in full; in case they could recognize the original recording from which a given extract was taken (which is mostly relevant for the music conditions), participants were instructed to judge similarities based on the sole extract, and not on the unpresented original material.

2. Categorization task

In the second task, each of the 12 stimuli of the tested condition is presented again one after the other. For instance, after judging similarities in a set of 44 triads of SM, a participant is asked to categorize each of the individual SM

TABLE I. Influence of splicing manipulation on inter-subject agreement on most-agreed value for categorization task.

		Normal condition	Spliced condition	Significance
Music	Source	$\beta_{NM}=0.79$	$\beta^*_{SM}=0.59$	$F(1,80)=18.23,\ p<0.001$
	Mood	$\beta_{NM}=0.67$	$\beta^*_{SM}=0.37$	$F(1,62)=73.05,\ p<0.001$
	Type	$\beta_{NM}=0.64$	$\beta^*_{SM}=0.36$	$F(1,8)=10.56,\ p=0.012$
Soundscapes	Source	$\beta_{NS}=0.89$	$\beta^*_{SS}=0.52$	$F(1,56)=30.19,\ p<0.001$
	Mood	$\beta_{NS}=0.68$	$\beta^*_{SS}=0.38$	$F(1,32)=28.70,\ p<0.001$
	Type	$\beta_{NS}=0.75$	$\beta^*_{SS}=0.46$	$F(1,20)=6.75,\ p=0.017$

recordings that composed the triads. For each stimulus, users were instructed to check through a list of descriptions that they think could apply (forced choice). As before, participants were given no other information than the sole audio presentation of the stimuli. The presented descriptions were organized in three broad categories (sources, moods, and types), which content varied depending on the tested condition (full list given in the Appendix).

Each category was designed to contain descriptions that apply naturally for all items (e.g., "car" sounds for a "boulevard" soundscape) as well as distractors, which do not apply[14] for any of the presented data (e.g., "violin" source for music and "boat" sounds for soundscapes). In each category, a wildcard description ("I don't know") was also included. Apart from the wildcard choice, none of the descriptions were enforced to be mutually exclusive, i.e., participants were left free to check as many as they thought could apply. Prior to the test, participants received the following instructions:

> This time, the task may require you to think analytically about the signals: Which instruments/sound sources you perceive, how you judge the sound emotionally, how you would describe the genre of the music, or to which type of environment the recording corresponds?

Additionally, when tested in one of the spliced conditions (SM and SS), participants were instructed to do the following:

> [...] please try to describe what you think the signal was *before* scrambling. For instance, all scrambled signals tend to sound rhythmically complex and maybe evoke tensed emotions associated with randomness. Please try to refrain from judging such properties, but rather focus on what you can perceive of the original signal. Imagine you are listening to the sound over a very strangely distorted telephone: We want you to describe the sound played at the other end, not to describe the telephone.

As before, each successive stimulus could be listened to as many times as needed; answers could be given before listening to a stimulus in full; participants were instructed to categorize the content of the sole extracts, rather than any unpresented context they might recognize.

D. Apparatus

Data were gathered in four sessions held on 4 days of December 2007, with randomized conditions. All participants were tested in the same room, in a calm environment, with the same apparatus: Audio was presented with closed-type headphones (always the same model), and the test application (a set of php scripts running on a local server) ran on G4 Apple computers. All instructions and tests were conducted in Japanese. Participants were instructed to set the audio volume to a comfortable level before the test. Three experimenters were present at any time to give instructions and answer queries. Participants were rewarded 1000 JPY (cf. 10 USD) for their participation, which took between 30 and 60 min.

III. RESULTS

A. Influence of splicing on categorization

In this section, we measure how much the descriptions (sources, moods, and types) that apply consensually to the normal stimuli can still be judged consistently in the corresponding spliced stimuli.

For a given stimulus s in a given condition c, we define the "agreement on best value" for a given description d $\beta_c(s;d)$ as the proportion of participants who agree on the most-agreed value (i.e., true or false) of description d for s. For example, when an extract of a *Beethoven* piano sonata is presented, most participants agree that the value for the description *piano* (a member of the *source* category) is "true." For stimuli s in a *spliced* condition c (i.e., one of SM and SS), we then compute the "agreement on best normal value" $\beta^*_c(s;d)$ as the proportion of participants (among all participants tested in condition c) who agree on the most-agreed value (i.e., true or false) of description d judged for the corresponding normal stimuli. We then average over all stimuli s in condition c, then all descriptions d in a given category (i.e., one of "source," "mood," and "type"). With the same example as above, $\beta^*_{SM}(s;d)$ measures the proportion of participants judging description *piano* to be true for the spliced variant of the sonata.

We observe the following:

- There is a severe and significant degradation of the agreement measured in spliced conditions over the most-agreed values judged in the corresponding normal conditions, when considering only *true* values in normal condition,

i.e., increased number of false negatives (see Table I).
- No significant degradation of agreement due to splicing when considering only *false* values in normal condition, i.e., no increase in false positives.
- Splicing reduces the number of hits per (stimulus and description) pair (participants check less descriptions for spliced signals than for the corresponding signals in normal condition), for all category groups and for both music and soundscape, but only significantly so for soundscape sources [number of checks: $M=2.75<M=1.89$, $F(1,19)=7.93$, $p=0.011$].
- The degradation of β_c^* due to splicing on the categorization tasks is of similar degree for music and soundscapes, for the mood category [$\Delta=-0.3$, but $F(1,47)=0.02$, $p=0.891$] and the type category [$\Delta=-0.28$, but $F(1,14)=0.01$, $p=0.944$]. The degradation of the agreement on source values is more severe for soundscapes than music: The agreement difference from NM to SM $\beta^*_{SM}(s;d\in\text{source})-\beta_{NM}(s;d\in\text{source})$ ($M=-0.20$) is smaller than from NS to SS $\beta^*_{SS}(s;d\in\text{source})-\beta_{NS}(s;d\in\text{source})$ ($M=-0.37$), with statistical significance [$F(1,68)=7.48$, $p=0.008$].

These results show that splicing significantly hinders the recognition of sources, moods, and types associated with stimuli in their normal conditions. This masking effect is not significantly different for music and soundscapes for both mood and type descriptions, but more important for soundscape sources than for music sources.

B. Influence of splicing on similarity performance

In this section, we measure how much of the similarity relations established by participants listening to normal stimuli ("ground truth") can still be identified by the participants listening to their spliced variants.

For a given triad t in a condition c, we define the "agreement on best pair" $\alpha_c(t)$ as the proportion of participants who agree on the most-agreed pair of "most" similar stimuli. For example, when presented a triad composed of two Beethoven piano sonatas and a *Coltrane* jazz piece, most participants judge that the pair of most similar stimuli is the one composed of the two sonatas. For triads t in the spliced conditions c, we then compute the "agreement on best *normal* pair" $\alpha_c^*(t)$ as the proportion of participants who agree over the most consensual pair judged by the participants tested on the same triad, but in the normal condition. We then average over all triads in the condition. With the same example as above, $\alpha^*_{SM}(t)$ is the proportion of participants who, when presented the triad composed of the spliced variants of the two sonatas and the jazz piece, judge that the pair of most similar stimuli is the one composed of the two spliced sonatas.

We observe the following:

- There is a significant degradation of agreement in the spliced conditions over the pairs most agreed on in the normal conditions, i.e., if two stimuli are consensually judged similar to one another in their normal condition, their spliced variants are not necessarily judged similar. This is true[15] for music ($[\alpha_{NM}]=0.86>\alpha^*_{SM}=0.53$, $F(1,48)=6.78$, $p=0.012$] as well as soundscapes [$\alpha_{NS}=0.91>[\alpha^*_{SS}]=0.77$, $F(1,48)=6.28$, $p=0.016$].
- However, the degradation due to splicing is 2.5 times more severe for music than for soundscapes: The agreement difference $\alpha_{NM}(t)-\alpha^*_{SM}(t)$ ($M=0.33$, SD=0.27) is larger than $\alpha_{NS}(t)-\alpha^*_{SS}(t)$, ($M=0.13$, SD=0.25) with statistical significance [$F(1,44)=6.38$, $p=0.015$].

SSs have only slightly different similarity relations to one another than in the normal condition, whereas the relations consensually judged in NM are lost when it is spliced. This opposes all effects found on categorization, which is equally hindered for music and soundscapes (and even more for soundscapes in the case of "source" categories).

IV. DISCUSSION

We find that splicing significantly impedes the recognition of source, mood, and type descriptions that can be consensually made (by our tested participants) for stimuli in their normal condition. We also find that splicing has a significant, but small influence on similarity performance for soundscape signals; for music, however, the similarity relations that are consensually established for normal signals are completely lost by splicing.

The difference observed between music and soundscape signals when spliced may result from several undesired effects. First, similarity judgments could be more difficult (i.e., less consensual) for music than for soundscapes even in the normal condition. However, we found that similarity judgments for NM signals are as consistent across participants as judgments for NS signals [$\alpha_{NM}=0.70<\alpha_{NS}=0.77$, but $F(1,86)=3.34$, $p=0.071$], so this effect is ruled out.

Second, judging the similarity of SM may be a more disorienting task than for SSs. But again, we found that similarity judgments for spliced signals are as consistent as judgments for normal signals ($\alpha_{NM}=0.70>\alpha_{SM}=0.66$ [$F(1,86)=1.27$, $p=0.262$]; $\alpha_{NS}=0.77>\alpha_{SS}=0.76$ [$F(1,86)=0.13$, $p=0.719$]). Poor similarity precision on SM is *consistently* observed for all participants.

Third, our results may be explained by the participants' lack of training for the task of listening to spliced signals. While a positive effect of training cannot be ruled out, it is interesting to note that judgments made on spliced signals reach the same degree of consensus than judgments made on ecologically valid, normal signals. For instance, the categories judged for spliced stimuli are as consistent as the categories for normal signals, e.g., for music moods [$\beta_{NM}(\text{mood})=0.90<\beta_{SM}(\text{mood})=0.90$, $F(1,26)=0.03$, $p=0.873$], types [$\beta_{NM}(\text{type})=0.85<\beta_{SM}(\text{type})=0.85$, $F(1,22)=0.00$, $p=0.998$], or sources [$\beta_{NM}(\text{source})=0.86<\beta_{SM}(\text{source})=0.83$, $F(1,29)=0.35$, $p=0.561$].

Finally, the difference observed between music and soundscapes may result from uncontrolled recognition effects of familiar music. If participants recognize the music stimuli used in the experiment, they may judge music similarity in the normal condition based on extra-acoustical information such as the artists' country. Since this recognition

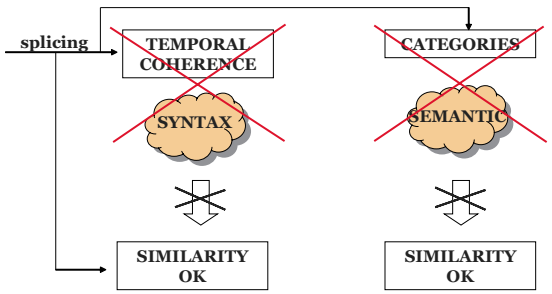


FIG. 1. (Color online) A schematic representation of the links found here between similarity, syntax, and semantics for soundscape perception. Splicing affects the temporal coherence of the audio signals, which was found to hinder activation in brain centers responsible for the processing of linguistic syntax. Since splicing preserves soundscape similarity to a large extent, we find here that soundscape similarity does not logically require syntactic processing of the type involved by speech and music (left). Moreover, splicing is found to hinder categorization. Therefore, soundscape similarity does not require the identification of the categories usually elicited by normal stimuli, most notably sound sources (right). See main text for details.

is likely lost when spliced, this information cannot be recruited to compare SM, and performance is degraded. For soundscapes, however, this recognition effect is negligible, which could explain the lesser difference from normal to spliced. This explanation can probably be ruled out, too. First, we found in debriefing the experiments that the participants did not recognize the stimuli in the NM condition (pieces that are relatively well known in the western world are not necessarily well known in Japan). Second, it was recently reported that listeners can make a distinction between familiar and unfamiliar music even in spliced stimuli with frame size between 50 and 100 ms.[16]

A. Splicing and syntax

Spliced signals sound distinctively unnatural. Physically, splicing preserves long-term statistical distributions of frames as well as local spectral content (since each individual frame is left untouched), but loses short-term dynamical relationships between frames. In particular, the frame rate adopted here is typically faster than the note rate in music, so even intra-note consistency (transient, sustain, and decay) is not preserved. Brain imaging studies show that spliced musical signals (for frame size around 500 ms) engender decreased activation in the left inferior frontal cortex,[17,18] a region closely associated with the processing of linguistic structure, or *syntax*, in spoken and signed languages.

We found here that humans are capable of comparing soundscapes in a way which resists splicing: Two soundscapes deemed similar to one another in normal condition are likely to be also found similar when listened in spliced condition. This suggests that judging the similarity of soundscapes does not require syntactic processing of the type involved by speech and music (Fig. 1, left). Whether syntactic processing is activated by the perception of non-SSs but contributes little, or whether it is simply not recruited even in normal situations, is left open for further investigation.

B. Splicing and semantics

Our results confirm that splicing is a categorization mask for non-speech audio signals. Degraded agreements for "true" categorical values, but not for "false" values, show that splicing masks the recognition of normally perceived categories (increased number of false negatives), but does not introduce any confabulated categories (no significant increase in false positives): One may not hear guitar in a spliced guitar signal, but one is unlikely to hear piano instead. This effect is the same for music and soundscapes.

These results are consistent with the use of splicing for speech signals,[11] and can be easily explained by the well-documented importance of the temporal envelopes for, say, musical instrument recognition.[5] However, they appear to contradict studies focusing on the perception of very short sounds. In a 1999 study (published only recently), Gjerdingen and Perrott[19] observed that, given as little as 250 ms of audio, human listeners were able to recognize the genre of a musical extract correctly more than 70% of the times. Similar effects were reproduced for further categories (including moods) for extracts as small as 50 ms,[20] which is the size of the frames we use here. That this ability should not hold for *sequences* of frames is intriguing and could point to an important role of short-term memory and attentional processes.

From our data, it is difficult to clarify the causal relation between loss of temporal coherence (by construction) and distorted ability to categorize. Friend and Farrar[21] studied the effect of splicing speech on judgments of the speaker's affective state, and found that it increased ratings of anger and excitement. They suggested that this depends on the speech-mimetic features of the resulting time series. Spliced signals contain regular artificial transients (due to the arbitrary concatenation of non-adjacent frames) and give an impression of high tempo and limited pausing regardless of the original signal. These features are often correlates of anger in normal speech. For music and soundscapes, we observed here that splicing typically did not introduce confabulated categories, and we could not find any systematic effect of splicing on the categorization of mood descriptions. However, similar effects cannot be ruled out and should be further investigated.

C. Soundscape similarity does not require categorization

The better similarity results obtained in spliced condition for soundscapes than for music could be explained by a possible recognition effect, which would be retained in the case of soundscapes. If more analytical details (i.e., descriptions) were preserved in SSs than in SM, then their comparably better similarity could be based on a comparison of high-level features in the way proposed by Tversky.[2] However, we found that splicing degrades categorization to a similar degree for both music and soundscape, and even to a greater degree for soundscapes when considering sound sources, so this effect can be ruled out. The fact that splicing preserves more of the music sources (instruments) than of the soundscape sources would in effect predict better similarity performance for music than for soundscapes, if based on analytical features. Moreover, we find here that relatively

good soundscape similarity does not require source, mood, or type categorization (we have the former without the latter in the spliced condition) (Fig. 1, right).

First, this suggests that a significant part of the process of comparing two soundscapes does not require high-level semantic features, but rather relatively raw representations, arguably generic. Second, this result indicates, in particular, that humans are able to give accurate and consensual judgments of the similarity of soundscapes with no prior recognition of their constituent sound sources. This is a surprising result, in apparent contrast with previous studies, which usually conclude that perceptive judgments of categorization and similarity on soundscapes are based on the identification, and then connotation, of sources. Psycholinguistic studies[9] only report on correlations. There are well-studied effects of "*a posteriori* rationalization" in such methodologies,[22] in which participants produce a lexical explanation ("because there are cars") for a cognitive process (here, "these two soundscapes are similar"), which may be independent from these causes.

We should be careful in concluding too strongly on "similarity without categorization." It is probably impossible to prevent categorization in normally cognitive participants, even for highly distorted stimuli such as the ones used here. What we are losing here by splicing is referential categorization in the conceptual system associated with normal stimuli: Participants find it impossible to map a spliced stimulus to categories that apply for normal stimuli, such as "piano" or "dog bark." However, when presented a spliced stimulus, participants likely build categories in a *ad hoc* semantic system adapted to spliced sounds (e.g., "sounds that go *flpflpflp*"). What we find here is that these *ad hoc* categories bear no relation with the "natural" categories: This spontaneous categorization of spliced signals does not help the participants retrieve the categories used for normal conditions. It is possible yet, if properly trained with pairs of normal and spliced stimuli, that participants can learn a mapping between both semantic systems, e.g., learn that "sounds that go *flpflpflp*" are usually spliced variants of songs containing *guitar*.

D. Does music similarity require categorization?

We found here that severely degraded music similarity (−33% agreement) is correlated with severely degraded categorization (cf. −30% agreement). From these data, we cannot conclude on the causality between these two processes. First, maybe similarity requires categorization, i.e., two music stimuli that are said to "sound the same" can only be said so because they are first categorized with a variety of descriptors (e.g., musical instruments, moods, etc.) and then found to share a significant number of such descriptors. If this is so, then splicing acts as a content mask on categorization and impedes similarity by hurting its prerequisite. Conversely, maybe categorization requires similarity, i.e., music classification operates as an exemplar matching process[23] based on a partly generic assessment of similarity (e.g., if it sounds like a rock song, then it is rock). The content-masking property of splicing could then be the consequence, rather than the cause, of this degraded similarity. Finally, maybe music categorization and similarity operate in conditional independence given a common cause, such as syntactic processing, which is degraded by splicing. This should be submitted to further investigation, e.g., with the help of brain imaging techniques, which are increasingly used to characterize semantic processing both for music[24] and natural sounds.[25]

ACKNOWLEDGMENTS

The authors would like to thank Brian Gygi, Geraint Wiggins, and anonymous reviewers for their useful comments. This research was partially funded by a JSPS Postdoctoral Fellowship for Foreign Researcher. Yusuke Kanayama helped with the logistics of the experimentation and Masumi Nasukawa with the Japanese translation of our instructions.

APPENDIX: DESCRIPTIONS USED IN CATEGORIZATION TASK

The following are descriptions for NM and SM:

- Sources: acoustic guitar, bass, brass (trumpet, etc.), drums, electric guitar (saturated), electric guitar (clear), harmonica, piano, strings (violin, etc.), synthesizer, vocals (male), vocals (female), and wind (flute, etc.).
- Moods: aggressive, calm, carefree, cerebral, cold, energetic, fun, happy, intimate, party/celebratory, passionate, plaintive, sad, and tense/anxious.
- Types: blues, club/dance, classic rock and roll, contemporary rock, country, electronic, experimental, folk, hard rock, heavy metal, jazz, pop, punk, and psychedelic.

The following are descriptions for NS and SS:

- Sources: car, truck, bus, bird, voice (male), voice (female), voice (children), boat, motorbike, footsteps, plane, train, industrial machinery, building site equipment, dog bark, water noise (rain, sea), collision noise, and car horns/siren.
- Moods: calm, disturbing, familiar, noisy, pleasant, restful, stimulating, stressful, unbearable, and unpleasant.
- Types: airport, avenue, calm street, construction site, factory, highway/motorway, intercity road, outway for emergency vehicles, park, railway, school yard, shopping/pedestrian street, street market, and technical equipment.

[1]R. M. Nosofsky, "Attention, similarity and the identification-categorization relationship," J. Exp. Psychol. Gen. **115**, 39–57 (1986).
[2]A. Tversky, "Features of similarity," Psychol. Rev. **84**, 327–352 (1977).
[3]R. L. Goldstone and J. Son, *Similarity* (Cambridge University Press, Cambridge, 2005), pp. 13–36.
[4]G. L. Murphy and D. L. Medin, "The role of theories in conceptual coherence," Psychol. Rev. **92**, 289–316 (1985).
[5]S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psychol. Res. **58**, 177–192 (1995).
[6]T. Eerola, T. Jarvinen, J. Louhivuori, and P. Toiviainen, "Statistical features and perceived similarity of folk melodies," Music Percept. **18**, 275–296 (2001).
[7]J. Ballas, "Common factors in the identification of an assortment of brief everyday sounds," J. Exp. Psychol. Hum. Percept. Perform. **19**, 250–267 (1993).

[8]V. Peltonen, A. Eronen, M. Parviainen, and A. Klapuri, "Recognition of everyday auditory scenes: Potentials, latencies and cues," in Proceedings of the 110th Convention of the Audio Engineering Society, Amsterdam, The Netherlands (2001).
[9]D. Dubois, C. Guastavino, R. Maffiolo, and M. Raimbault, "A cognitive approach to soundscape research," J. Acoust. Soc. Am. **115**, 2495 (2004).
[10]J.-J. Aucouturier, B. Defreville, and F. Pachet, "The bag-of-frame approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music," J. Acoust. Soc. Am. **122**, 881–891 (2007).
[11]K. R. Scherer, "Randomized splicing: A note on a simple technique for masking speech content," Journal of Experimental Research in Personality **5**, 155–159 (1971).
[12]M. Castellengo and D. Dubois, "Is timbre a property of the signal, of the instrument, or a cognitive construct?," in Proceedings of the 2005 Conference on Interdisciplinary Musicology (CIM05), Montreal, Canada (2005).
[13]With this instruction, we aim to direct participants toward holistic rather than selective listening—these two strategies have been found to induce significantly different brain signatures, see P. Janata, B. Tillmann, and J. Bharucha, "Listening to polyphonic music recruits domain-general attention and working memory circuits," Cogn. Affect. Behav. Neurosci. **2**, 121–140 (2002).
[14]Or at least were not expected to apply consensually.
[15]The values reported here discard outlier triads with particularly low consensus: seven triads for condition NM and four triads for condition NS.
[16]B. Tillmann and E. Bigand, "Global context effects in normal and scrambled musical sequences," J. Exp. Psychol. Hum. Percept. Perform. **27**, 1185–1196 (2001).
[17]D. J. Levitin and V. Menon, "Musical structure is processed in 'language' areas of the brain: A possible role for brodmann area 47 in temporal coherence," Neuroimage **20**, 2142–2152 (2003).
[18]T. Matsui, K. Kazai, M. Tsuzaki, and H. Katayose, "Investigation of the musician's brain activation during different music listening modes: A near-infrared spectroscopy study," in Proceedings of the 2008 International Conference on Music Perception and Cognition, Sapporo, Japan (2008).
[19]R. Gjerdingen and D. Perrott, "Scanning the dial: The rapid recognition of music genres," J. New Music Res. **37**(2), 93-100 (2008).
[20]R. Ashley, "Affective and perceptual responses to very brief musical stimuli," in Proceedings of the 2008 International Conference on Music Perception and Cognition, Sapporo, Japan (2008).
[21]M. Friend and M. J. Farrar, "A comparison of content-masking procedures for obtaining judgments of discrete affective states," J. Acoust. Soc. Am. **96**, 1283–1290 (1994).
[22]P. Johansson, L. Hall, S. Sikström, B. Tärning, and A. Lind, "How something can be said about telling more than we can know: On choice blindness and introspection," Conscious Cogn **15**, 673–692 (2006).
[23]R. M. Nosofsky, "Exemplar-based accounts of relations between classification, recognition and typicality," J. Exp. Psychol. Learn. Mem. Cogn. **14**, 700–708 (1988).
[24]S. Koelsch, E. Kasper, D. Sammler, K. Schulze, T. Gunter, and A. Friederici, "Music, language and meaning: Brain signatures of semantic processing," Nat. Neurosci. **7**, 302–307 (2004).
[25]A. Cummings, R. Ceponiene, A. Koyama, A. P. Saygin, J. Townsend, and F. Dick, "Auditory semantic networks for words and natural sounds," Brain Res. **1115**, 92–107 (2006).

The masking of interaural delays

Andrew J. Kolarik and John F. Culling[a)]
School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, United Kingdom

(Received 5 December 2006; revised 26 January 2009; accepted 28 January 2009)

Four experiments measured discrimination of interaural time delay (ITD) thresholds for broadband noise in the presence of masking noise of the same bandwidth as the target (0.1–3 kHz for experiments 1–3 and 0–10 kHz for experiment 4). In experiments 1–3, listeners performed interaural two–interval two–alternative forced–choice (2I-2AFC) delay discrimination tasks with stimuli composed of delayed and masking noises mixed in proportions of delayed noise ranging between 1 and 0.05. Experiments 1–3 employed interaurally correlated, anticorrelated, and uncorrelated maskers, respectively. Experiment 4 measured centering accuracy for continuous noise with a range of interaural coherences (equivalent to proportion of delayed noise) obtained by mixing delayed and interaurally uncorrelated noises. Results indicate that in the presence of an interaurally correlated masker ITD thresholds doubled for every halving of the proportion of delayed noise power in the stimulus. This function became steeper as the masking noise changed from interaurally correlated, to uncorrelated, to anticorrelated. The results were compared to thresholds predicted by a model based on variations in the distribution of interaural phase differences of the stimulus components.
 [DOI: 10.1121/1.3083238]

PACS number(s): 43.66.Ba, 43.66.Dc, 43.66.Mk, 43.66.Pn [RLF] Pages: 2162–2171

I. INTRODUCTION

Listeners can use three primary acoustic cues for sound localization: interaural time differences (ITDs), interaural level differences (ILDs), and spectral cues introduced by the head, torso, and pinnae. At low frequencies, listeners tend to rely on the ITD cue for sound lateralization (Wightman and Kistler, 1992). At higher frequencies, listeners tend to rely on ILDs to lateralize sounds (Kuhn, 1977). Elevation and disambiguation between the frontal and rear hemifields are provided by monaural spectral changes in the sound caused by reflections among the corrugations of the pinna (Lopez-Poveda and Meddis, 1996), head movements that change the binaural cues (Perret and Noble, 1997), and movements of the sound source that the listener controls (Wightman and Kistler, 1999). In spite of these cues, the presence of masking noise has a debilitating effect on listeners' abilities to localize sounds. In order to quantify these effects, a range of studies has investigated localization (Jeffress *et al.*, 1962; Good and Gilkey, 1996) and lateralization (Egan and Benson, 1966; Cohen, 1981; Ito *et al.*, 1982) in the context of masking noise. The effect of masking noise in the presence of delayed noise is also important when modeling binaural temporal windows using a lateralization paradigm (Wagner, 1991; Bernstein *et al.*, 2001).

Previous research has demonstrated the adverse effects of masking noise on ITD discrimination, and that different types of masking noise vary in their impact on listeners' abilities to lateralize narrow-band signals. Specifically, lateralization of a signal embedded in masking noise is least disrupted by the presence of interaurally correlated noise and most disrupted by anticorrelated noise, with disruption from uncorrelated noise falling between the two.

Egan and Benson (1966) investigated listener's abilities to lateralize a monaural tone with a duration of 250 ms at frequencies of 500 and 390 Hz embedded in binaural noise (either interaurally correlated or uncorrelated). Lateralization thresholds were obtained by varying the signal-to-noise ratio (SNR) adaptively using a 2I-2AFC task. In the presence of interaurally correlated noise, lateralization thresholds were found to be 4–5 dB lower than when an uncorrelated masker was present.

Cohen (1981) measured ITD just-noticeable-differences (jnds) around zero ITD for a 250 Hz sinusoid presented against broadband masking noise which was interaurally correlated, uncorrelated, or anticorrelated. The ITD jnd was measured as a function of SNR. For a constant SNR, jnds were highest for the anticorrelated masking condition and lowest when the masker was interaurally correlated. jnds were intermediate when the masker was uncorrelated. The slope of the thresholds versus SNR was steeper when the masker was uncorrelated than when the masker was interaurally correlated, and the two slopes converged at approximately 19 dB above N0S0 detection threshold (denoted MoSo in Fig. 2 of Cohen, 1981). Thus, at high SNRs, ITD jnds were at their lowest and equal for both interaurally correlated and uncorrelated masking noises. At lower SNRs, jnds were larger when the masker was interaurally uncorrelated. When the masker was interaurally anticorrelated, ITD jnds were larger than jnds for interaurally correlated and uncorrelated masking noises, and increased as SNR was reduced.

Finally, Ito *et al.* (1982) measured interaural time jnds for narrow-band noise (1/3 octave wide) in the presence of a broadband masker. jnds were measured at a fixed SNR for a range of interaural configurations of the masker: interaurally correlated, anticorrelated, and uncorrelated, plus quiet (no masker), and six conditions that manipulated the perceived

a)Author to whom correspondence should be addressed.

lateral position of the masker. Similar to Cohen's (1981) study using tonal stimuli, jnds were smallest for the interaurally correlated masking condition, larger for the uncorrelated masker, and largest for the anticorrelated masking condition. No consistent relation was observed between the perceived lateral position of the masker and the magnitude of the jnd. Ito *et al.* (1982) were able to successfully model the hierarchy of ITD jnds for each interaural configuration of the masker using a model based on variations in the distribution of interaural phase differences of the stimulus components.

This effect of masking noise on ITD processing has been used in studies that have investigated the temporal resolution of the binaural system. Temporal resolution can be modeled using a temporal window that is described by one or more time constants, the duration of which describes the resolution of the system. In the studies of Wagner (1991) and Bernstein *et al.* (2001), temporal windows were measured for barn owls and human participants, respectively, by presenting a burst of interaurally delayed noise (the probe) temporally contiguous with masking noise. It was assumed that the listener detects the ITD imposed on the probe by centering a temporal window at the midpoint of the probe in order for the maximum amount of interaurally delayed noise to fall within the window. The window integrates together the instantaneous interaural delays of the probe with the interaural delays of the masking noise that also falls within the window. In order to model a temporal window, it is assumed that the integration results in an internal, effective ITD that is lower than the external ITD imposed on the probe (Bernstein *et al.*, 2001). Given that the probe and surrounding noise have equal power, the integration should yield the same internal ITD value as direct mixing at the same energy ratio. As a result, the external ITD must be increased to a magnitude that will bring the internal ITD up to threshold. When the probe duration is long, ITD thresholds will be low as very little masking noise enters the window, reducing the internal effective ITD. As the probe duration decreases, more masking noise enters the window and is integrated with the probe ITD, increasing ITD threshold.

It is important to quantify the effect of the masking noise on ITD threshold, as this is explicitly modeled during the window fitting procedure. Best-fit ITD thresholds are calculated by using a coefficient that relates threshold to proportion of delayed noise present. Bernstein *et al.* (2001) effectively assumed that when a broadband interaurally delayed noise target (N_τ) is heard in the presence of masking interaurally correlated or uncorrelated noise (N_0 or N_u), the effective internal ITD is doubled for each halving of the proportion of delayed noise present within the window; that is, that the masking noise dilutes the interaurally delayed noise. Having fit windows using the dilution assumption in the case of interaurally correlated masking noise, Bernstein *et al.* (2001) assessed how well the same windows fitted data in the case when the masking noise was interaurally uncorrelated. They were modeled using the same assumption, but with an adjustment to account for a change in sensitivity, as thresholds in the presence of interaurally uncorrelated masking noise were higher than corresponding thresholds when the masker was interaurally correlated. Experiments 1 and 3 investigate whether the dilution assumption is valid for interaurally correlated and uncorrelated masking noises, respectively.

In the latter case, the dilution assumption appears to run contrary to the findings of Jeffress *et al.* (1962), who measured participants' centering judgments at a range of interaural correlations by mixing interaurally correlated and uncorrelated noises. The data of Jeffress *et al.* (1962) show that for each halving of the coherence (which is equivalent to the proportion of delayed noise present), thresholds are substantially *less than* doubled (see Figs. 4 and 5, bottom right panel).

The following experiments investigate the effects of masking interaurally correlated (experiment 1), anticorrelated (experiment 2), and uncorrelated (experiments 3 and 4) noises on participants' ability to lateralize broadband noise using the localization cue provided by ITD, by directly mixing the target and masking noise in various proportions.[1] The aims of these experiments were fourfold.

The first aim was to test the dilution assumption in the case of interaurally correlated, anticorrelated, and uncorrelated maskers; when a broadband interaurally delayed noise target (N_τ) is heard in the presence of masking noise, does the detectable ITD double for each halving of the proportion of delayed noise present?

The second aim was to compare listeners' abilities to lateralize a signal embedded in noise to previous studies. The results of Cohen (1981) and Ito *et al.* (1982) found the smallest ITD jnds for interaurally correlated masking conditions, larger jnds for uncorrelated maskers, and the largest jnds for anticorrelated masking conditions. If the results of the current experiments follow this ordinal relationship among masking noises, then ITD threshold slope as a function of proportion of delayed noise present for interaurally uncorrelated masking noise will fall between the anticorrelated and the correlated masking noise slopes. Results with an interaurally uncorrelated masker can be directly compared to those of Jeffress *et al.* (1962), but comparisons with Egan and Benson (1966), Cohen (1981), and Ito *et al.* (1982) are complicated by the difference in bandwidth of the delayed signal.

The third aim was to quantify the effects of masking noise in order to apply the findings to the modeling techniques employed when measuring the binaural temporal window using a lateralization paradigm. This was achieved by measuring masking coefficients (the slope describing ITD threshold versus proportion of delayed masking noise) for interaurally correlated, anticorrelated, and uncorrelated masking noises.

The fourth aim was to compare the results of the experiments to thresholds predicted by the model of Ito *et al.* (1982), which is based on interaural phase difference.

II. METHODS COMMON TO EXPERIMENTS 1–3

The stimuli were generated digitally with a sampling rate of 44.1 kHz and 16-bit sample depth using MATLAB, band-limited between 100 and 3000 Hz prior to presentation, and gated with 10-ms raised-cosine onset and offset ramps. Listeners were presented with the stimuli over Sennheiser

HD590 headphones at an overall sound level of 75 dB (A), played through a 24-bit Edirol UA-20 sound card and passed through an MTR HPA-2 headphone amplifier in a single-walled Industrial Acoustics Company (IAC) sound-attenuating booth within a sound-deadened room. Trial-by-trial feedback was provided. Interaural time delays were imposed on the stimuli by adding a ramp function to the phase spectrum at one ear, resulting in an ongoing ITD (i.e., no onset-time difference).

III. EXPERIMENT 1: INTERAURALLY CORRELATED MASKING NOISE

In the first experiment, participants were given a 2I-2AFC discrimination task, where noise with an interaural correlation of 1 was mixed with interaurally delayed noise so that each interval was composed of delayed noise and correlated noise mixed at a predetermined proportion. Half the delay was imposed on interval 1, and half the delay, in the opposite direction, was imposed on interval 2. Consistent with the model of Bernstein *et al.* (2001), it was hypothesized that for each halving of the proportion of delayed noise present, the threshold ITD would double. Thus, on a log-log plot, according to the dilution assumption, as the proportion of delayed noise decreased and more interaurally correlated noise was mixed into the stimuli, thresholds would increase in a linear manner on a 1:1 slope. The slope describing ITD threshold versus proportion of delayed interaurally correlated noise was defined here as the correlated masking coefficient (CMC).

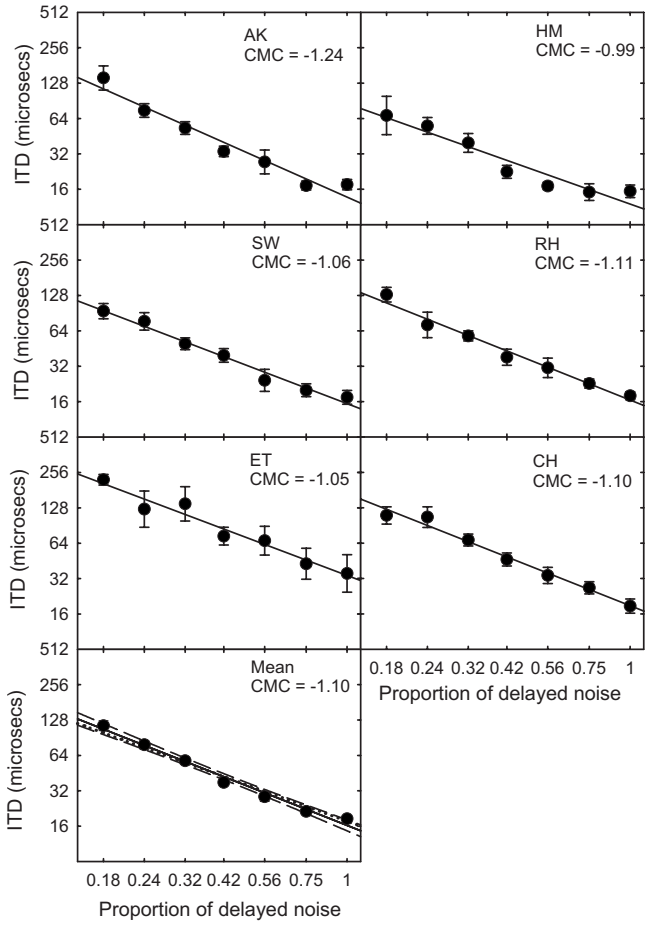


FIG. 1. Results for experiment 1 (correlated masker). Solid regression lines are plotted for individual listeners and the mean. The dotted line represents a slope of −1. Dashed lines represent 95% confidence intervals. Both axes are plotted logarithmically.

A. Stimuli and procedure

Six participants (including one of the authors) took part in the experiment. One was male and five were female, aged between 18 and 25. All participants except for ET and RH had previous experience with psychophysical experiments. Untrained participants received at least 5 h of training before data collection. Excluding the author, they were paid upon completion.

Listeners' performed a two-interval discrimination task. Three independent Gaussian noises were generated (N_1 and N_3), and a delayed copy of $N_1(N_2)$. The magnitude of the delay imposed on N_1 was half that of the delay difference between the two intervals. To keep the power constant, the noises were scaled in amplitude in the following manner, such that p is the proportion of the total power of the stimulus that is made up of delayed noise.

$$\text{Left channel: } \sqrt{p}N_1 + \sqrt{1-p}N_3. \quad (1)$$

$$\text{Right channel: } \sqrt{p}N_2 + \sqrt{1-p}N_3. \quad (2)$$

The second interval was created using the same procedure, except that the delayed noise was presented to the opposite ear. The adapted variable was the difference in ITD between the two intervals. The inter-stimulus interval was 500 ms, and each interval had a duration of 100 ms. ITD thresholds were obtained for $p=1$, 0.75, 0.56, 0.42, 0.32, 0.24, and 0.18.

The listener's task was to identify in which direction the sound image moved from interval to interval (left to right or right to left), corresponding to the ITDs embedded within each interval. Participants pressed "1" if the noise moved from right to left, and "2" if it moved left to right. The sign of the interaural delay was randomized and trial-by-trial feedback was provided. The ITD was varied adaptively in order to obtain a 70.7% correct estimate (Levitt, 1971). Each adaptive track started with the ITD set to 500 μs. Initially, the step size of the adaptive track corresponded to a factor of 0.2 and was reduced to a factor of 0.05 following two reversals. Threshold was obtained after a further ten reversals. The last ten reversals composed a measurement phase, and the average of the reversals within the measurement phase was taken as threshold.

Each listener performed nine experimental runs. Measurement began after performance had stabilized; experimental thresholds were taken by averaging the remaining runs after the listener achieved thresholds below 500 μs at each proportion of delayed noise (nine runs for AK, CH, RH, and SW, eight runs for HM, and four runs for ET).

B. Results

Figure 1 shows threshold ITDs measured at different proportions of delayed noise mixed with interaurally correlated noise. When $p=1$, all the noise within each interval was delayed, producing the lowest thresholds for the six listeners (an average of 18.5 μs). Threshold ITD was found to approximately double for each halving of delayed noise present. The slope of the regression line of the mean observed data was -1.10 and accounted for 98% of the variance. The dotted line represents predicted values based on a CMC of -1 and falls within the 95% confidence intervals of the mean data, showing that the data do not deviate significantly from this trend.

C. Discussion

As the proportion of delayed noise decreased and more interaurally correlated masking noise was mixed into the intervals, thresholds increased in a log-linear manner on an approximately 1:1 slope so that for each halving of the delayed noise present, the threshold ITD doubled, supporting the dilution assumption of Bernstein *et al.* (2001).

The following two experiments characterized the effect of interaurally anticorrelated (experiment 2) and uncorrelated (experiment 3) masking noises on ITD threshold and examined whether thresholds change in the same manner as they do in the presence of interaurally correlated masking noise. However, we modified the method for these experiments compared to experiment 1 because we anticipated a potential problem. If one considers cross-correlation as a mathematical approximation to human ITD processing, the effect of increasing ITD will be to move a peak in the cross-correlation function further and further from zero. As ITD increases, the contrast between the cross-correlation functions for a positive and a negative ITD increases monotonically as the peaks separate. However, as there must be some limit to the range of delays over which cross-correlation is performed by the auditory system, this contrast eventually saturates and begins to decline, suggesting that discrimination performance will decline once again. Perceptually, the laterality reaches an extreme at somewhere around the ecological limit of naturally occurring ITDs, and then images start to become diffuse (Blodgett *et al.*, 1956). Adding interaurally uncorrelated noise can only be expected to make the task harder, as the noise will reduce the size of the cross-correlation peaks without altering their locations. Cohen (1981) and Ito *et al.* (1982) found that both interaurally uncorrelated and anticorrelated masking noises produced higher ITD jnds than correlated masking noise, raising the danger that, for a low proportion of delayed noise, threshold might be close to a nonmonotonic section of the underlying psychometric function. With these facts in mind, we decided to obtain psychometric functions in experiments 2 and 3, rather than the adaptive track paradigm in order to avoid potential violation of assumption 1 of Levitt (1971) (p. 468): "The expected proportion of positive responses is a monotonic function of stimulus level (at least over the region in which observations are made)."

To obtain a psychometric function, stimuli were produced with a range of fixed ITDs, and the percentage of correct responses at each ITD was recorded. From these data, the 71% correct point of the psychometric function can be interpolated from a fitted function to obtain threshold values that can be directly compared to those obtained using the adaptive track method.

IV. EXPERIMENT 2: INTERAURALLY ANTICORRELATED MASKING NOISE

The log-linear slope describing ITD threshold versus the proportion of interaurally anticorrelated noise was defined here as the anticorrelated masking coefficient (AMC). This was measured in experiment 2.

A. Stimuli and procedure

Three participants (including one of the authors) took part in the experiment. One was male and two were female, aged between 18 and 28. All participants had previous experience with psychophysical experiments. Excluding the author, they were paid upon completion.

Listeners' performed a 2I-2AFC discrimination task. Stimuli were generated in the same way as in experiment 1, but with interaurally anticorrelated (correlation$=-1$) rather than correlated masking noise. That is to say, using Eqs. (1) and (2) except that N_3 was inverted in one channel. Intervals were presented so that the delayed noise in the first interval was presented to the ear favored by the interaural delay embedded in the interval, and the delayed noise in the second interval presented to the opposing ear. The inter-stimulus interval was 500 ms.

The listener's task was to identify in which direction the sound image moved between the two intervals (left to right or right to left), corresponding to the ITDs embedded within each interval. The direction of ITD change was randomized and trial-by-trial feedback was provided. The participants were tested at a stimulus duration of 100 ms with ITDs of 1024, 512, 256, 128, 64, 32, 16, and 8 μs. Initially, a block of trials was presented with proportions of 1, 0.5, 0.2, and 0.1, the order in which the eight ITDs were presented within the block was randomized, and 20 trials were repeated for each ITD. A second block of trials with proportions of 0.75, 0.25, 0.15, and 0.05 followed. Each block consisted of 8 ITDs$\times$4 proportions$\times$20 repetitions$=$640 trials in all. Following two training runs, participants performed three experimental runs at each proportion. The average of the three runs was taken as threshold. Two-parameter (slope and threshold) logistic functions were fitted to the data, and the 71% points of the functions were taken as thresholds in order to compare the thresholds with those taken in experiment 1 using the adaptive track procedure.

B. Results

Figure 2 shows threshold ITDs measured at different proportions of delayed noise mixed with interaurally anticorrelated noise. When $p=1$, all the noise within each interval was delayed, producing the lowest thresholds for the three listeners. Although listener EO's thresholds were substan-

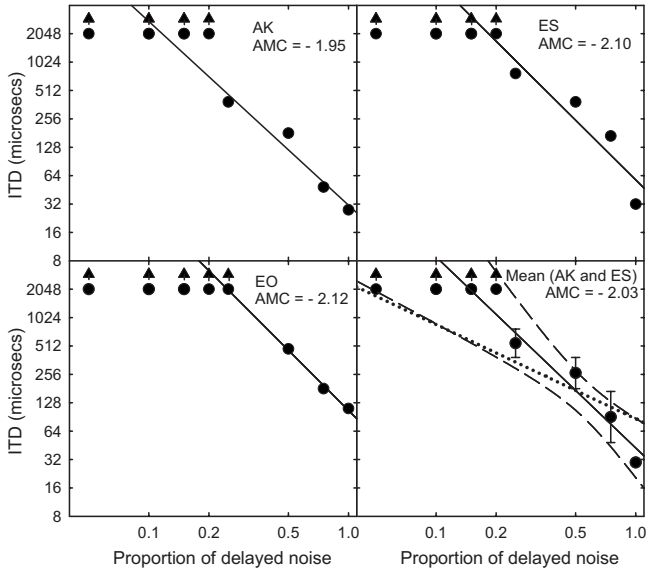


FIG. 2. Results for the experiment 2 (interaurally anticorrelated masker). Solid regression lines are plotted for individual listeners and the mean. Mean thresholds are plotted as an average across data for listeners AK and ES only, as listener EO displayed lower sensitivity than these listeners. Upward pointing arrows indicate that a measure of threshold could not be obtained for those coherences as the 71% point of the psychometric function fell outside the measured range. The dotted line represents a slope of −1. Dashed lines represent 95% confidence intervals. Both axes are plotted logarithmically.

tially higher than those of other listeners, the slope relating threshold to proportion of delayed noise present was approximately equal to that of other listeners, indicating that the sensitivity of this listener was lower than that of the others.[2] Consequently, mean thresholds in the bottom right panel of Fig. 2 are plotted as an average of listeners AK and ES only. Threshold ITD was found to more than double for each halving of delayed noise present. The slope of the regression line of the mean observed data was −2.03 and accounted for 92% of the variance. The dotted line represents predicted values based on a slope of −1 and falls outside the 95% confidence intervals of the mean data (represented by dashed lines in Fig. 2), showing that the slope of the data is significantly steeper than −1.

C. Discussion

As the proportion of delayed noise decreased and more interaurally anticorrelated noise was mixed into the intervals, thresholds increased in a log-linear manner on a slope of −2.03 so that for each halving of the delayed noise present, the threshold ITD more than doubled. As the slope of ITD thresholds was steeper than that observed when an interaurally correlated masker was employed (i.e., −1, see Fig. 1), the presence of interaurally anticorrelated noise was more disruptive to listener's abilities to lateralize a signal embedded in noise than interaurally correlated masking noise. These results follow the same hierarchy as those of Cohen (1981) and Ito *et al.* (1982), who found that ITD jnds were larger for interaurally anticorrelated than correlated masking conditions.

V. EXPERIMENT 3: INTERAURALLY UNCORRELATED MASKING NOISE

In experiment 3, the effect of interaurally uncorrelated masking noise on ITD threshold was assessed using stimuli comparable to experiments 1 and 2. The interaural coherence of each interval was manipulated by mixing interaurally delayed and uncorrelated noises. Note that when interaurally delayed and uncorrelated noises are added, the interaural coherence is equal to the proportion of delayed noise power in the stimulus, so the proportion of delayed noise is numerically equivalent to the coherence and to the interaural correlation when the noise delay is zero. The slope describing ITD threshold versus the proportion of interaurally uncorrelated noise was defined here as the uncorrelated masking coefficient (UMC).

Results obtained using an interaurally uncorrelated masker can be directly compared with those of Jeffress *et al.* (1962), who investigated the effect of interaural correlation on the ability of participants to center a noise. In that study, interaurally correlated and uncorrelated noises were mixed in each ear to obtain interaural correlations of 1, 0.75, 0.5, 0.25, 0.2, 0.15, 0.1, and 0. Participants adjusted a delay line in the presence of continuous noise in order to center the noise in their heads. The standard deviation of each participant's centering judgment was measured as a function of interaural correlation. This standard deviation should, in principle, be equivalent to an ITD threshold. The slope relating such a threshold ITD to correlation/coherence/proportion of delayed noise was −0.48 (see Fig. 4, bottom right panel). This slope is considerably shallower than the CMC values observed in experiment 1 with interaurally correlated masking noise that ranged from −0.99 to −1.24 (see Fig. 1).

To examine any effects of stimulus duration in experiment 3, participants were tested at three stimulus durations (100 ms, 500 ms, and 1 s). The long-duration stimulus conditions were included in order to be more comparable with the experiment of Jeffress *et al.* (1962) in which continuous noise was used.

A. Stimuli and procedure

Three listeners took part in experiment 3. One (the first author) was male and two were female, aged between 18 and 25. All listeners had participated in experiment 1. Excluding the author, they were paid upon completion.

Listeners' performed a 2I-2AFC discrimination task. To manipulate coherence, two independent Gaussian noises (N_1 and N_2) were generated and mixed: the left channel was presented containing noise N_1, and the right channel contained a mixture of two noises N_1 (delayed) and N_2 mixed in the proportion (p).

$$\text{Left channel:} N_1. \tag{3}$$

$$\text{Right channel:} pN_1 + \sqrt{1-p^2}N_2. \tag{4}$$

An interaural delay was added to one channel after noise mixing. Participants were tested at stimulus durations of 100 ms, 500 ms, and 1 s. Following two training runs, participants performed three experimental runs for each coherence

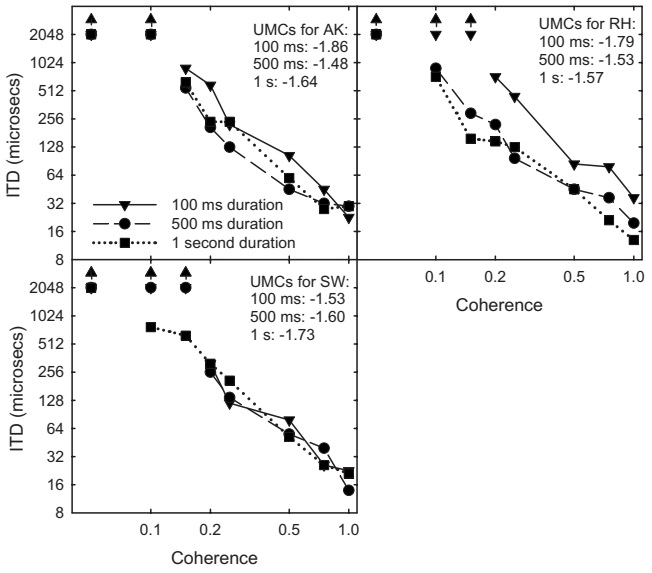


FIG. 3. Individual data for experiment 3 (interaurally uncorrelated masker) for stimulus durations of 100 ms (downward pointing filled triangles joined by solid lines), 500 ms (filled circles joined by dashed lines), and 1 s (filled squares joined by dotted lines). Thresholds represent 71% discriminability. Upward pointing arrows indicate that a measure of threshold could not be obtained for those coherences as the 71% point of the psychometric function fell outside the measured range. Both axes are plotted logarithmically.

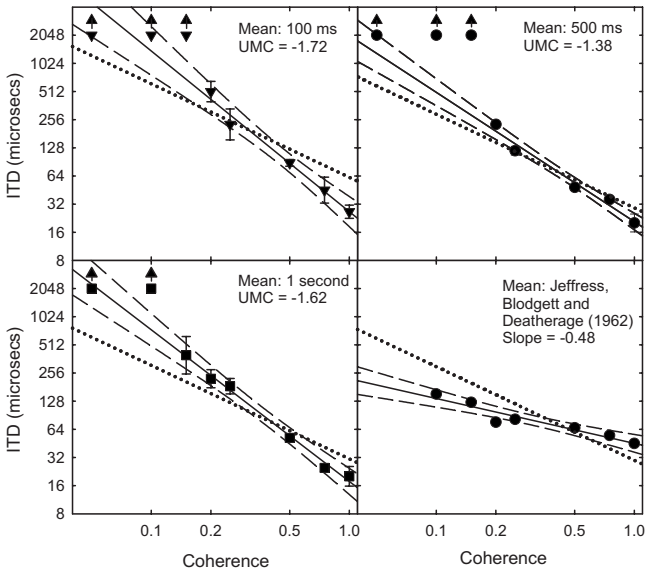


FIG. 4. Mean results for experiment 3 (interaurally uncorrelated masker). Each panel shows results at each of the three stimuli durations, and the bottom right panel shows mean data replotted from Jeffress *et al.* (1962), who used continuous stimuli. Solid regression lines are plotted for each duration. The plotted thresholds are 71% points averaged from the last three experimental runs. The dotted line represents thresholds with a slope of −1. Dashed lines represent 95% confidence intervals. Both axes are plotted logarithmically.

at each stimulus duration. Participants were tested in blocks of coherences of 1, 0.5, 0.2, 0.1, 0.75, 0.25, 0.15, and 0.05 by measuring psychometric functions using the procedure described in experiment 2.

B. Results

Figure 3 shows individual data for the three stimulus durations. UMCs ranged from −1.48 to −1.86 (for clarity, regression lines representing the UMC for each participant and duration are not plotted). The mean of the UMCs measured for individual participants across all listeners and stimulus durations was −1.64. When the coherence was unity, all the noise within each interval was delayed, producing the lowest thresholds for the three listeners. Threshold ITD was found to more than double for each halving of coherence. Thresholds at the lowest coherences could not be measured as the 71% point of the psychometric function fell outside the measurable range.

Figure 4 shows the mean of the data across listeners for the three stimulus durations. The dotted lines represent predicted values based on a UMC of −1. At all three stimulus durations the dotted lines fall outside of the 95% confidence intervals of the regression (represented by dashed lines in Fig. 4), in each case indicating that the slope of the data was significantly steeper than −1. The mean UMCs for each duration ranged from −1.38 to −1.72, which were substantially higher than the slope of −0.48 obtained by Jeffress *et al.* (1962). No consistent effect of stimulus duration was observed, and the range of UMCs obtained across stimulus duration and listener was wide (−1.48 to −1.86).

C. Discussion

As coherence decreased, thresholds rose so that each time the coherence was halved, thresholds more than doubled (see Figs. 3 and 4). Once again, this finding contrasts with the effect on ITD of interaurally correlated masking noise observed in experiment 1, where thresholds were observed to double, and indicates that the dilution hypothesis does not hold for the case where an interaurally uncorrelated masker is present.

An explanation of the differing influences of interaurally correlated, anticorrelated, and uncorrelated masking noises was put forward by Ito *et al.* (1982) based on their model. They measured interaural time jnds for narrow-band noise in the presence of a broadband masker. They were able to model the hierarchy of ITD jnds at a fixed SNR of approximately +10 dB for various interaural configurations of the masker (smallest for the interaurally correlated masker, larger for the uncorrelated masker, and largest for the anticorrelated masker). Their model was based on variations in the distribution of interaural phase differences of the stimulus components and similar to the models proposed by Webster (1951) and Jeffress *et al.* (1956).

The explanation of Ito *et al.* (1982) essentially encompassed signal detection notions of how changing the ratio between the mean (μ) and the standard deviation (σ) is predicted to affect sensitivity, that discriminability depends on the differences in the means relative to the standard deviation of the two ITD distributions. When no masking noise was present, thresholds were approximately equal for all types of masking noise (see Figs. 1–4), but when masking noise was added the distribution of ITDs became wider and the standard deviation increased, requiring an increase in the magni-

tude of ITD to support discrimination. According to this model, the slopes of the CMC, UMC, and AMC differ because, as Ito *et al.* (1982) showed, the variability of short-term estimates of ITD about their mean of zero was comparably small in the presence of an interaurally correlated masker, larger when an uncorrelated masker was present, and largest when an anticorrelated masker was present. The predictions of the model of Ito *et al.* (1982) are evaluated more fully in Sec. VI.

Steeper UMCs at all three stimulus durations were observed in the current experiment compared to the slope observed by Jeffress *et al.* (1962) (see Fig. 4). These results suggest that the shallow slope obtained in their experiment was probably not due to their use of continuous stimuli, as the UMC obtained with each stimulus duration was considerably steeper than the slope obtained by Jeffress *et al.* (1962). The following experiment was carried out in order to replicate the experiment of Jeffress *et al.* (1962), and explored the possibility that the shallow slope of their results stemmed from the narrow range of ITDs that were employed in their study.

VI. EXPERIMENT 4: REPLICATION OF Jeffress *et al.* (1962)

As described in the introduction to experiment 3, Jeffress *et al.* (1962) asked listeners to center a noise in their heads by means of adjusting a delay line with an ITD range of ± 450 μs. It is not clear whether an adjustment beyond the ITD limit resulted in the ITD reaching a hard limit where the delay line reached its maximum and the knob consequently stopped moving, or whether the sign of the interaural delay was reversed at this point, flipping the lateralization of the stimulus from one side of the head to the other. It seems reasonable to assume that the latter was the case[3] as the former would give the listener a rather obvious cue. Even so, it is possible that an unintended cue was still present. The distance from the control knob's "flipping point" when the ITD limit was reached could have been used by the listener as a method of producing relatively accurate centering. Increasing the range of the ITDs available would reduce the usefulness of this cue because the laterality produced by ITD decreases above about 1 ms (Blodgett *et al.*, 1956) which would prevent the stimulus "flipping" in perceived laterality. This possibility was addressed in the following replication of the experiment of Jeffress *et al.* (1962), where participants performed the task with ITDs ranging from ± 450 μs, 1 ms, and 2 ms. An increase in threshold slope with increasing ITD range would support the hypothesis that a restricted range provided a cue based on the stimulus flipping in laterality when the ITD limit was reached.

A. Stimuli and procedure

Three listeners took part in experiment 4. One (the first author) was male and two were female, aged between 18 and 25. The author had participated in experiments 1–3, and the others were naïve participants. Excluding the author, they were paid upon completion.

The stimuli consisted of continuous broadband (0–10 kHz) noise generated by concatenating 102.4-ms segments of noise drawn at random from a 5-s buffer. The coherence of the stimuli was manipulated using the same mixing method implemented in experiments 1 and 2, but with N_3 replaced by an independent noise at one ear. All noises were obtained from the 5-s buffer. The participants' task was to center the noise in their heads by using keyboard controls. Pressing '1' resulted in a decrement in stimulus ITD of 36 μs, '2' a decrement of 9 μs, "3" an increment of 9 μs, and "4" an increment of 36 μs. Pressing "x" terminated the program. If the ITD limit was reached, the sign of the interaural delay was reversed. Each stimulus trial was initially presented with a random ITD within the specified range. Initially, participants were presented with a block of trials with coherences of 1, 0.75, 0.5, 0.25, 0.2, 0.15, 0.1, and 0 with an ITD range of ± 450 μs in order to familiarize them with the experimental setup. Following this training, participants were given another block of trials with the ITD ranging from ± 450 μs, and blocks of ± 1 and ± 2 ms followed. This procedure was repeated three times, and the standard deviation of the three runs at each ITD range was taken as threshold.

All stimuli were generated on-line at a sampling frequency of 20 kHz using a TDT AP2 array-processor card and presented via a TDT System-2 psychoacoustic rig (DD1, PA4, FT5-9, and HB6) and presented over Sennheiser HD650 headphones at an overall sound level of 70 dB (A) in a single-walled IAC sound-attenuating booth within a sound-deadened room. Filtering was performed using an FT6 reconstruction filter with a low pass cut-off frequency of 10 kHz. No feedback was given.

B. Results

Figure 5 shows results from the replication of Jeffress *et al.* (1962) at ITD ranges of ± 0.45, ± 1, and ± 2 ms. The dotted lines represent predicted values based on a UMC of -1. The slope of the data at an ITD range of ± 0.45 ms (-0.54) was comparable to the slope obtained by Jeffress *et al.* (1962), which was -0.48. As the range increased, the slope of the data became steeper. At a range of ± 2 ms, the slope of the data (-1.67) was comparable to the UMCs observed in experiment 3.

C. Discussion

The data from experiment 4 show that the shallow slope of ITD thresholds obtained by Jeffress *et al.* (1962) was probably due to the limited range of ITDs (± 0.45 ms) that were employed in that study. The increase in threshold slope with increasing ITD range supports the hypothesis that a restricted ITD range provides a cue based on the stimulus flipping in laterality when the ITD limit was reached. Increasing the range progressively reduced the usefulness of this cue because the laterality produced by ITD decreases above about 1 ms (Blodgett *et al.*, 1956).

VII. PREDICTIONS OF THE MODEL OF Ito *et al.* (1982)

The model developed by Ito *et al.* (1982) was implemented computationally in order to compare the experimen-

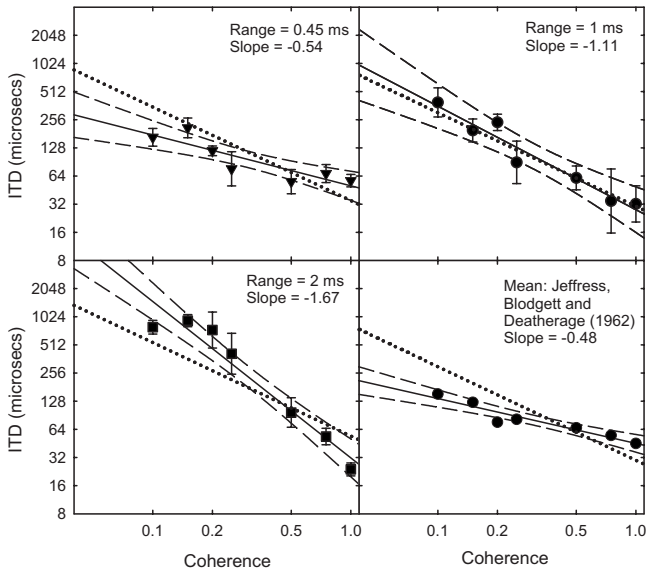


FIG. 5. Mean results for experiment 4. Each panel shows results with continuous stimuli at ITD ranges of ±0.45 μs [the range employed by Jeffress *et al.* (1962)], ±1 ms, and ±2 ms. Solid regression lines are plotted for each range of ITD. The plotted thresholds are averaged standard deviations for the three participants. The dotted line represents thresholds with a slope of −1. Dashed lines represent 95% confidence intervals. Both axes are plotted logarithmically.

tal results with thresholds produced by the model (see Appendix). Figure 6 (right panel) shows the effect of p on the predicted thresholds produced by our implementation of the model (each data point is the mean of four simulations) and compares it to the mean coefficients observed for experiments 1–3 (left panel). Interestingly, the ordinal relationship among the three types of masking noise changes as a function of p. For p below about 0.5 (corresponding to 0-dB SNR), interaurally uncorrelated noise was predicted to be the most effective masker. The fitted slope to the thresholds simulated for an interaurally correlated masker (−1.07) was very similar to the CMC obtained in experiment 1 (−1.10).

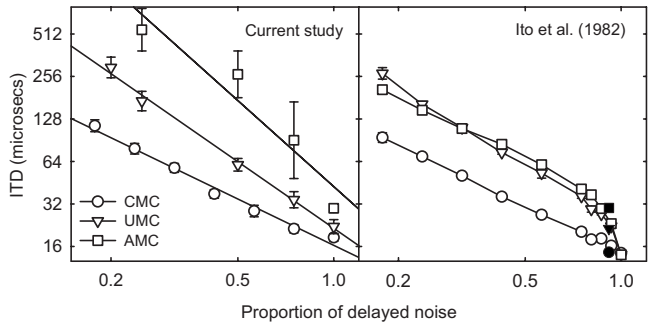


FIG. 6. Mean masking thresholds from the current study (left panel) and predicted thresholds produced by implementation of the model developed by Ito *et al.* (1982) (right panel). Circles indicate an interaurally correlated masker, triangles indicate an uncorrelated masker, and squares an anticorrelated masker. Solid regression lines are plotted for coefficients produced in the current study. The plotted CMC thresholds are averaged across all listeners, and the AMC thresholds are averaged over listeners AK and ES. Plotted UMC thresholds are averaged across all listeners and all stimuli durations. Thresholds plotted from the model of Ito *et al.* (1982) are averaged across four simulations. Filled symbols are the plotted means of thresholds taken from Fig. 1 of Ito *et al.* (1982). Both axes are plotted logarithmically.

The fitted slope for an interaurally uncorrelated masker (−1.52) was very similar to the mean UMC observed in the current study (−1.57), but the fitted slope for an anticorrelated masker (−1.35) was considerably shallower than the mean AMC observed in experiment 2 (−2.03). Thresholds simulated for masking interaurally uncorrelated and anticorrelated noises were observed to be curvilinear in the region of high correlation (see right panel of Fig. 6). Filled symbols indicate thresholds taken from Fig. 1 of Ito *et al.* (1982) for the first set of four listeners in interaurally correlated, uncorrelated, and anticorrelated conditions at a single SNR of +10.3 dB, equivalent to a proportion of delayed noise (p) of 0.915. Despite differences in the stimuli employed [the stimuli of Ito *et al.* (1982) were band-limited to 1/3 octave around 500 Hz (445–561 Hz) with a duration of 350 ms, whereas stimuli in the current experiment were broadband (0–3 kHz) with durations ranging between 100 ms and 1 s], the ordinal relation between thresholds obtained with interaurally correlated, uncorrelated, and anticorrelated masking noises and the magnitude of the ITD thresholds is similar across the two studies.

In summary, implementation of the model of Ito *et al.* (1982) produced a slope of thresholds for interaurally correlated masking noise that closely matched the CMC observed in experiment 1, and produced the ordinal relationship among the three types of masking noises for p values above approximately 0.5. However, the functions relating proportion of delayed noise to ITD threshold for interaurally uncorrelated and anticorrelated noises were incorrectly predicted to be curvilinear, and below $p=0.5$ the ordinal relationship changed, with uncorrelated noise predicted to produce more effective masking than anticorrelated noise.

VIII. GENERAL DISCUSSION

The results of the experiments described in this study indicated that the presence of masking noise resulted in an increase in the ITD threshold of delayed noise. In the case of an interaurally correlated masker, for each halving of the proportion of delayed noise present, threshold ITD doubled (i.e., ITD threshold slope=−1). ITD threshold slope was steeper than −1 in the case of an interaurally uncorrelated masker (−1.64), and steeper still in the case of an interaurally anticorrelated masker (−2.03). Thus, the experiments described in this paper indicate that masking noise with an interaural correlation of 1, anticorrelated noise, and uncorrelated noise would be integrated within a binaural temporal window in different ways. Modeling the temporal window should thus depend on the type of masker incorporated in the design of the experiment.

For example, Bernstein *et al.* (2001) measured temporal windows by presenting the listener with a burst of interaurally delayed noise (the probe) temporally fringed with masking noise. Best-fit thresholds were calculated in the following manner. First, the integral of the area occupied by the probe was computed and divided by the total integral of the window. This value represented the proportion of delayed noise within the window and is affected by the duration and the assumed shape of the window. Best-fit ITD thresholds

were then calculated by using a coefficient that related threshold to proportion of delayed noise present. By using the assumption that masking noise diluted the delayed noise, Bernstein *et al.* (2001) effectively determined predicted threshold ITDs using a coefficient of −1. The experiments described in this paper have allowed temporal windows to be measured using stimuli of a similar design to Bernstein *et al.* (2001) employing uncorrelated masking noise and determining predicted threshold ITDs using the UMC (Kolarik and Culling, 2009). When interaurally correlated masking noise is present, then the CMC (−1) should be modeled to predict thresholds, for uncorrelated masking noise the UMC (−1.64) should be modeled, and for anticorrelated masking noise the AMC (−2.03) should be modeled.

IX. CONCLUSIONS

The data described within this paper demonstrate the following conclusions.

(1) When interaurally correlated masking noise is mixed with delayed noise, ITD threshold doubles for each halving of delayed noise present.
(2) When interaurally anticorrelated or uncorrelated masking noise is mixed with delayed noise, ITD threshold more than doubles for each halving of delayed noise present. The presence of interaurally anticorrelated noise is more debilitating to the listener than the presence of uncorrelated noise.

ACKNOWLEDGMENTS

The work was supported by the EPSRC. Thanks are due to Steve Colburn and an anonymous reviewer for their helpful comments relating to earlier drafts of the manuscript. We would also like to thank Mike Akeroyd for his discerning advice and to Torsten Marquardt for his thoughtful comments. Summaries of these findings were presented at the September 2005 BSA and May 2006 ASA meetings.

APPENDIX

The model of Ito *et al.* (1982) of sensitivity to interaural delay under masking was implemented computationally using MATLAB. Stimuli similar to those used by Ito *et al.* (1982) in their experiments were generated and band-limited to 1/3 octave around 500 Hz (445–561 Hz) with a duration of 350 ms. To create the stimulus, delayed and masking noises were added together in the appropriate ratio. Ito *et al.* (1982) employed a single SNR of +10.3 dB equivalent to a proportion of delayed noise (p) of 0.915. The mean interaural phase difference was derived by calculating the analytic signal and thus the instantaneous phase for each channel, subtracting the instantaneous phase of the left channel from the right, wrapping values outside the range $\pm\pi$ back into that range, and then taking the average across the entire 350-ms stimulus. We found that implementing internal noise as additive interaural phase noise with a standard deviation of 1.5 rad as independent samples yielded good predictions of the data of Ito *et al.* (1982). If the resulting mean instantaneous interaural phase was positive, the model reported the stimulus on the right, and otherwise on the left. Predictions of percent correct were produced by repeating these operations for 200 such stimuli. Predicted thresholds were derived by generating psychometric functions for seven noise delays logarithmically spaced between 10 and 316 μs and taking the 75% point from a fitted logistic function.

[1]It is assumed that the binaural system does not distinguish between the temporal differences between sequential presentation of noises (as per the stimuli presented by Bernstein *et al.* (2001) and simultaneously presented delayed target and masking noise (as in the current experiments). For example, the temporal sequence of ITDs in the experiment of Bernstein *et al.* (2001) with interaurally correlated masking noise is described by a series of zero values during presentation of the interaurally correlated noise preceding the probe, followed by a series of values at the test ITD, followed by a series of zero values during the interaurally correlated noise lagging the probe. In the current experiments where delayed and masking noises are directly mixed, the sequence of ITD values is always random due to the combination of the delayed and masking noises. When delayed noise is mixed with interaurally correlated masking noise, the interaction of the two statistically independent noises results in a random distribution of noises with a mean determined by the relative power of the noise components, where the mean is between zero and the ITD of the delayed noise. When the delayed target noise and the masking noise are mixed within a hypothetical temporal window, it is assumed that it is still the proportion of delayed noise within a window that dictates threshold ITD, regardless of how it is temporally distributed.

[2]Participant EO received an additional 4 h of data collection in order to establish whether further training would result in lower thresholds. As no further improvement was observed, and as this listener was already previously experienced in psychophysical experiments, it was concluded that she was less sensitive than the others.

[3]Jeffress *et al.* (1962) were able to accurately predict listeners' chance performance at a correlation of zero by assuming a rectangular chance distribution among the points of the switch that listeners used to respond in a condition that was included as check on the experimental method. Had the end-points of the delay line been marked by a fixed limit to the knob movement, listeners would have been able to at least avoid setting the knob close to this limit.

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.

Blodgett, H. C., Wilbanks, W. A., and Jeffress, L. A. (**1956**). "Effect of large interaural time differences upon the judgment of sidedness," J. Acoust. Soc. Am. **28**, 639–643.

Cohen, M. F. (**1981**). "Interaural time discrimination in noise," J. Acoust. Soc. Am. **70**, 1289–1293.

Egan, J. P., and Benson, W. (**1966**). "Lateralization of a weak signal presented with correlated and with uncorrelated noise," J. Acoust. Soc. Am. **40**, 20–26.

Good, M. D., and Gilkey, R. H. (**1996**). "Sound localization in noise: I. The effect of signal-to-noise ratio," J. Acoust. Soc. Am. **99**, 1108–1117.

Ito, Y., Colburn, H. S., and Thompson, C. L. (**1982**). "Masked discrimination of interaural time delays with narrow-band signal," J. Acoust. Soc. Am. **72**, 1821–1826.

Jeffress, L. A., Blodgett, H. C., and Deatherage, B. H. (**1962**). "Effect of interaural correlation on the precision of centering a noise," J. Acoust. Soc. Am. **34**, 1122–1123.

Jeffress, L. A., Blodgett, H. C., Sandel, T. T., and Wood, C. L. (**1956**). "Masking of tonal signals," J. Acoust. Soc. Am. **28**, 416–426.

Kolarik, A. J., and Culling, J. F. (**2009**). "Measurement of the binaural temporal window using a lateralization task," Hear. Res. **248**, 60–68.

Kuhn, G. F. (**1977**). "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157–167.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lopez-Poveda, E. A., and Meddis, R. (**1996**). "A physical model of sound diffraction and reflections in the human concha," J. Acoust. Soc. Am. **100**, 3248–3259.

Perret, S., and Noble, W. (**1997**). "The effect of head rotations on vertical plane sound localization," Percept. Psychophys. **59**, 1018–1026.
Wagner, H. (**1991**). "A temporal window for lateralization of interaural time differences in barn owls," J. Comp. Physiol. [A] **169**, 281–289.
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.
Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.
Wightman, F. L., and Kistler, D. J. (**1999**). "Resolution of front-back ambiguity in spatial hearing by listener and source movement," J. Acoust. Soc. Am. **105**, 2841.

Precursor effects on behavioral estimates of frequency selectivity and gain in forward masking

Skyler G. Jennings,[a] Elizabeth A. Strickland, and Michael G. Heinz[b]
Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana 47907

(Received 15 February 2008; revised 19 November 2008; accepted 13 January 2009)

The experiments presented in this paper explore the hypothesis that cochlear gain is reduced, in a frequency-specific manner, over the course of a sound (called a "precursor") which was designed to activate the medial olivo-cochlear reflex (MOCR). Psychophysical tuning curves (PTCs) and off-frequency growth of masking (GOM) functions were measured with two precursors. The on-frequency precursor condition, which was hypothesized to activate the MOCR at the signal frequency, produced a PTC with a lower best frequency in all subjects consistent with less gain. This same condition produced a GOM function with less gain and an elevated compression breakpoint. The data were analyzed with two models. The gain-reduction model, which assumed a change in the basilar membrane input-output function, was superior at predicting the data relative to a model of additivity of masking. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3081383]

PACS number(s): 43.66.Dc, 43.66.Mk, 43.66.Ba [MW] Pages: 2172–2181

I. INTRODUCTION

The auditory system, like other sensory systems, adjusts to stimuli in the environment. Some adjustments may protect the ear from harm (such as the stapedial reflex), while others may facilitate perception of a signal of interest. For example, Dean *et al.* (2005) showed that rate-level functions in the inferior colliculus adjust to the statistical characteristics of the input waveform intensity. Such an adjustment may facilitate perception by optimizing auditory dynamic range and avoiding the deleterious effects of saturation. In humans, it may be possible to study how the auditory system adjusts to sound through perceptual experiments involving auditory masking. For example, under certain conditions the detection of a short signal, presented simultaneously with a masker, improves when preceded by a sound (often called a "precursor") rather than silence. This improvement in detection threshold has been referred to as "overshoot" (Zwicker, 1965) or the "temporal effect" (Hicks and Bacon, 1992) and has been a source of study for over 40 years. For generality, the term precursor may be regarded as a separate sound or an extension of the masker's duration.

Overshoot may occur as a result of an adjustment in the auditory system over the course of the precursor. Although the hypothesized mechanisms of this adjustment are still a matter of debate, recent psychophysical experiments suggest that it may be related to a decrease in cochlear gain (von Klitzing and Kohlrausch, 1994; Strickland, 2001, 2004; Strickland and Krishnan, 2005; Strickland, 2008). Such a decrease in gain may be mediated by the medial olivo-cochlear reflex (MOCR) (Schmidt and Zwicker, 1991). Efferent fibers from the MOCR connect directly to the outer hair cells in the cochlea. These hair cells, which are associated with the cochlear amplifier, amplify soft sounds (i.e., provide "gain") and sharpen tuning (or "frequency selectivity") by feeding energy into the cochlea's basilar membrane (BM). In addition, they help determine the BM's input/output (I/O) function, which describes how much gain is applied to a stimulus of a given intensity. Animal studies report a frequency-specific reduction in BM gain and frequency selectivity when the MOCR is elicited (Guinan and Gifford, 1988; Cooper and Guinan, 2006). The current study explored whether similar reductions in frequency selectivity and gain could be observed in humans through perceptual experiments involving precursors in forward masking.

The hypothesis that gain reduction explains overshoot seems counterintuitive because for some overshoot conditions, frequency selectivity increases when a precursor is present (Bacon and Viemeister, 1985; Bacon and Moore, 1986; Kimberley *et al.*, 1989; Bacon *et al.*, 2002; Strickland, 2004). Although a broadband precursor has been shown to decrease frequency selectivity (Strickland, 2001), a notched precursor, having the same spectrum as the masker, increases frequency selectivity. Strickland (2004) hypothesized that suppression may account for this dichotomy. This hypothesis was supported by showing that the data were consistent with a BM I/O model which assumed gain decreased in cochlear regions overlapping with the frequency spectrum of the precursor. The author argued that at wide notch widths, suppressive masking largely determined thresholds and reasoned that suppression may decrease over the course of the precursor. If this were the case, masking due to suppression would also decrease. Such an effect would require higher masker levels at wide masker notch widths and produce higher estimates of frequency selectivity.

If the opposing findings regarding frequency selectivity in overshoot are related to a reduction in suppression, it may be possible to control for suppression by using a forward masking technique. Although precursor studies related to forward masking and frequency selectivity are limited, insight may be garnered from studies on forward masker duration. Unfortunately, these studies are inconclusive. For example,

[a]Electronic mail: sgjennin@purdue.edu
[b]Holds an additional appointment in the Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana 47907.

Kidd *et al.* (1984) reported a slight broadening of tuning as masker duration increased while Bacon and Jesteadt (1987) reported a slight sharpening of tuning when measuring psychophysical tuning curves (PTCs) and masking patterns (MPs), respectively. This failure to show a consistent effect may be understood by considering a conceptual model of forward masking that assumes two underlying masking mechanisms which differ in their time-course of influence. The first mechanism has a short time-course and masks the signal by virtue of a continuation of masker-related neural activity (Oxenham, 2001). This mechanism is theoretically similar to the temporal window model (Moore *et al.*, 1988) and is assumed to be effective for short maskers (or parts of longer maskers) existing 20–30 ms prior to the signal's onset (i.e., neural activity "persists" for at least 20–30 ms). The second mechanism, which has a longer time-course, produces masking by reducing the gain of the cochlear amplifier. This mechanism is assumed to be effective for maskers existing at least 40–70 ms prior to the signal's onset. Specifically, the masking effect of this second mechanism is assumed to follow the time-course of the MOCR as described by Backus and Guinan (2006). Under the assumption that forward masking is a mix of neural persistence and gain reduction, the PTC and MP data discussed above become difficult to interpret because neither masking mechanism is under experimental control. In other words, for a given masker frequency and duration, the relative contribution of each mechanism is unknown.

Krull and Strickland (2008) described a precursor-based technique that may control or isolate the masking effects of the gain-reduction mechanism. This technique took advantage of the "sluggishness" of the MOCR by restricting the masker duration and the masker-signal interval such that gain-reduction effects from the masker should be minimal (i.e., gain-reduction effects were primarily from the precursor). The study by Krull and Strickland (2008) was limited to estimating cochlear gain via growth of off-frequency masking; however, if the MOCR were involved, frequency selectivity may also change in a manner consistent with less cochlear-amplifier gain. The current set of experiments adopts the technique described by Krull and Strickland (2008) to test this hypothesis. Behavioral estimates of frequency selectivity and gain were measured using PTCs and off-frequency growth of masking (GOM) functions, respectively. Within each experiment, data were measured for two conditions which differed by the frequency of the precursor. In the on-frequency precursor condition, the precursor was the same frequency as the signal and was assumed to elicit the MOCR at the signal place. Conversely in the off-frequency precursor condition, the precursor's frequency was fixed well below the signal frequency and was assumed to have little to no effect on the cochlear amplifier at the signal place. Under the assumption of MOCR-induced gain reduction, the on-frequency precursor condition may result in a broader PTC and a GOM function consistent with less gain. In the general discussion, a model is described that tests the assumption of MOCR-induced gain reduction. An additional model based on additivity of masking is also tested.

II. METHODS

A. Subjects and procedures

Four normal-hearing subjects were recruited for the experiment. All were clinically normal on measures of acoustic immittance, distortion-product otoacoustic emissions, and pure-tone audiometry (thresholds were below 15 dB HL at audiometric frequencies between 250 and 4000 Hz). Subjects were between 24 and 27 years of age and were recruited among the students and staff at Purdue University and citizens of the Greater Lafayette community. All were inexperienced with psychoacoustic tasks and paid for their time except subject 3, who is the first author. Prior to data collection, the subjects were given 2–3 h of practice on each experiment to minimize learning effects.

In both of the experiments, masker level at threshold was measured using a three-interval forced-choice task. Subjects listened to the stimuli in a double-walled sound-attenuating booth. For a given trial, the listener pressed a button to indicate the interval in which the signal was perceived. Visual stimuli marked the observation intervals and feedback was provided to indicate a correct or an incorrect response. If the response was incorrect, the masker level decreased. If the response was correct over two successive trials, the masker level increased. This stepping rule converged on the level where the subject achieved 70.7% correct (Levitt, 1971). A run consisted of 50 trials. An even number of reversals, excluding the first two, were averaged to estimate the threshold for a run. The step size was 5 dB until the second reversal, after which it decreased to 2 dB. At least two and often three runs were averaged for the final threshold estimate of each condition. Runs with a standard deviation greater than 5 dB were excluded. If performance improved appreciably over successive runs, measurement continued until performance stabilized, after which only the last two or three measurements were averaged. Each experimental session was limited to 1–1.5 h in length.

B. Stimuli

The digitally-generated stimuli were presented through four separate D/A channels (TDT DA 3-4), low-pass filtered at 10 kHz (TDT FT5 and TDT FT6-2), and adjusted by programmable attenuators (TDT PA4). The stimuli were then mixed (TDT SM3), passed through a headphone buffer (TDT HB6), and presented to the listener's left ear via an ER-2 insert earphone.

The stimulus paradigm used for the GOM and PTC experiments is depicted in Fig. 1. This paradigm contains the following three components: (1) a fixed-level tonal precursor, (2) a variable-level tonal masker, and (3) a fixed-level 4 kHz signal. In all conditions, the total durations (including onset/offset ramps) of the precursor, masker, and signal were 160, 20, and 6 ms, respectively. Each had 5 ms $\cos^2$ onset and offset ramps except the signal, where this value was 3 ms. There was no delay between the offset of one component (i.e., precursor, masker, or signal) and the onset of the next. High-pass noise (cutoff frequency $= 1.2 \times$ signal frequency) was gated 50 ms before precursor onset and 50 ms after signal offset at a spectrum level 40 dB below the signal level

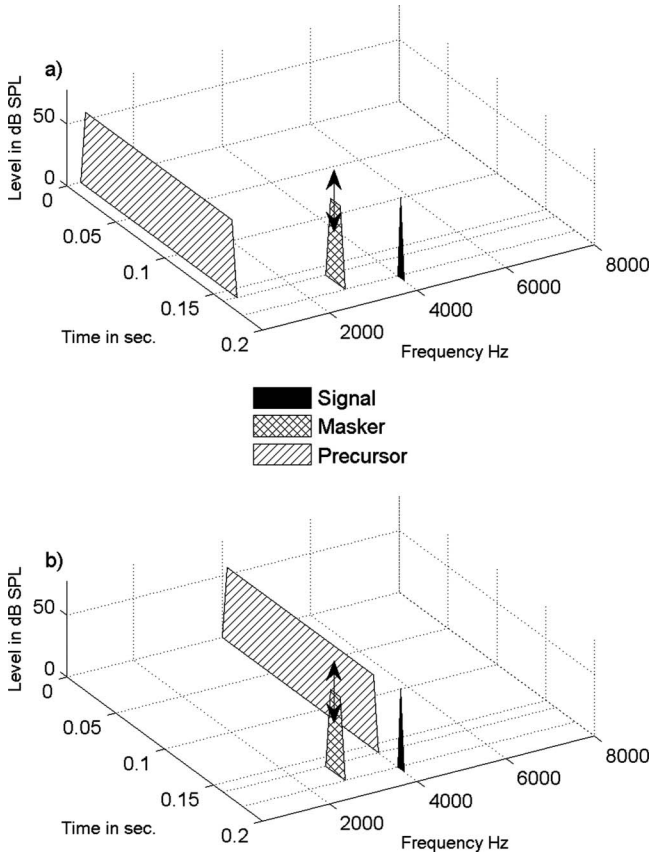


FIG. 1. Three dimensional representation of the stimuli used in the experiments. Temporal and spectral parameters of the signal and precursor are shown for the off-frequency (a) and on-frequency precursor (b) conditions. Masker level at threshold was the dependent variable in all conditions. In experiment 1, thresholds were measured for a series of masker frequencies (Note: only one frequency is displayed in the figure) to construct two PTCs, one for each precursor frequency. In experiment 2, thresholds for an off-frequency masker were measured for a series of signal levels to construct two GOM functions, one for each precursor frequency.

to restrict off-frequency listening. Within each experiment, data were collected for two precursor frequencies. The 4 kHz precursor ("on-frequency precursor") was designed to elicit the MOCR at the signal place, while the 0.8 kHz precursor ("off-frequency precursor") functioned as a control. This stimulus paradigm attempted to restrict MOCR-induced gain reduction to cochlear regions at and near the signal by (1) fixing the level, frequency, and duration of the on-frequency precursor across all conditions and (2) minimizing masker duration and masker-signal delay such that MOCR-induced gain reduction from the masker was minimal during the presentation of the signal.

In addition to attempting to control for the MOCR, this stimulus paradigm minimized any attention-related effects by including a precursor in all conditions rather than comparing measurements with and without a precursor. Data from a pilot experiment (Fig. 2) found that thresholds for the 0.8 kHz precursor condition were nearly identical to thresholds without a precursor. These pilot data are from S3 and involve a slightly longer masker (40 ms) than the experiments described below.

Thresholds for the signal were measured in quiet for all subjects and are displayed in Table I. Also displayed is the

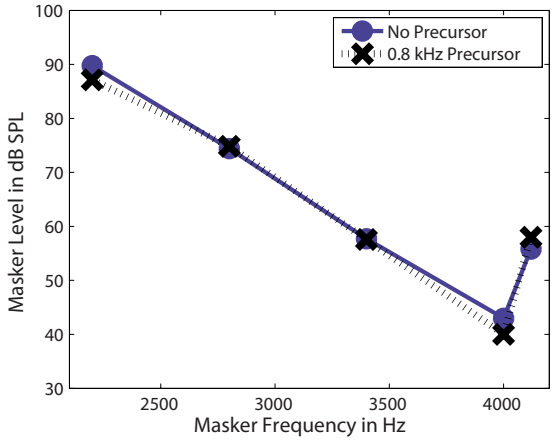


FIG. 2. (Color online) Data comparing PTCs measured without a precursor and with the 0.8 kHz control precursor. Note that these conditions produce similar thresholds and suggest the 0.8 kHz control precursor is an appropriate approximation of thresholds without a precursor.

level of the signal in the PTC experiment, the level of the precursor (L_{PE}), and the level of the precursor needed to just mask the signal in the absence of the masker (L_{PA}).

C. Experiment 1: Psychophysical tuning curves

Experiment 1 estimated frequency selectivity at the signal place. The signal level was fixed at 17–23 dB SL (i.e., approximately 20 dB SL). The masker level was adjusted to find threshold for a series of masker frequencies ranging from approximately 2000 to 5000 Hz. Data were collected in steps of approximately 400 and 80 Hz for masker frequencies below and above the signal frequency, respectively. The precursor level was fixed at 60 dB SPL for all subjects except subject 4, who was unable to perform the task at this level. For this subject, the precursor level was lowered to 50 dB SPL.

D. Experiment 2: Growth of masking functions

Experiment 2 estimated the I/O function at the signal frequency using the GOM technique first described by Oxenham and Plack (1997). This technique often involves comparing on and off-frequency GOM data to account for any non-linearities in forward masking (Oxenham and Plack, 2000; Oxenham and Bacon, 2004); however, for short masker-stimulus intervals (0–10 ms) the on-frequency GOM function is linear with a slope of 1. This suggests that the off-frequency GOM function can be directly interpreted as an estimate of the I/O function without measuring on-frequency GOM. The masker frequency was fixed at 2.8 kHz. A similar experiment (Oxenham and Plack, 1997) reported using a masker frequency one octave below the signal frequency. Initial data collection revealed that it was not possible to measure threshold for such a low masker frequency without exceeding the limits of the equipment (95 dB SPL). Therefore, the masker's frequency was raised sufficiently to measure threshold and maintain a reasonable assumption of linear growth at the signal place. Thresholds were obtained for a series of fixed signal levels from near threshold to 40–60 dB SL in 3 or 5 dB steps.

TABLE I. Threshold data and model parameter estimates.

Subject	Signal threshold in quiet[a]	Signal level (PTC)[a]	Precursor level (L_{PE}) [a]	Precursor masking level (L_{PA}) [a]
S1	27.30	50.00	60.00	70.58
S2	38.14	55.00	60.00	74.90
S3	35.08	52.00	60.00	94.00
S4	35.03	55.00	50.00	71.01

		Roex filter parameters								GOM parameters			
Subject	Precursor	BF (kHz)	L_{BF} [a]	pl	wl	tl	pu	rms[b]	Q_{10}	Gain[b]	c	BP1[a]	rms[b]
S1	Off-frequency	4.00	53.8	42.8	7.0×10^{-3}	12.2	>500	0.5	9.72	40	0.17	42.83	1.3
	On-frequency	3.53	40	33.7	4.6×10^{-3}	8.2	34.7	1.8	4.22	13.3	0.62	70	2.2
S2	Off-frequency	4.12	57.3	21.3	NA	NA	224.9	2.2	4.92	25.62	0.07	64.13	0.8
	On-frequency	3.74	50.4	25.8	NA	NA	27.1	0.6	3.36	⋯	0.34	56.75	0.3
S3	Off-frequency	4.08	57.0	26.0	NA	NA	>500	0.5	6.68	32.99	0.42	55	1.5
	On-frequency	3.92	47.1	30.0	NA	NA	110.7	1.7	5.89	19.32	0.55	67.42	3
S4	Off-frequency	4.10	48.3	26.0	NA	NA	>500	1.2	6.36	22.20	0.32	62.72	1.5
	On-frequency	3.91	39.3	33.5	1.0×10^{-4}	0.16	105.7	6.44	5.3	14.00	0.54	68.6	1.7

[a]Units in dB SPL.
[b]Values in dB.

III. RESULTS AND DISCUSSION

Figure 3 displays the individual data for the PTC and GOM experiments in the top and bottom rows, respectively. Squares and asterisks are data for the off-frequency and on-frequency precursors. Dashed and solid lines are model fits to the data and are discussed below.

A. PTC data

Masker level is plotted as a function of masker frequency for the PTC data. Error bars were omitted for clarity [the average standard deviations by subject number for the off- and on-frequency data were 2.38, 3.82 (S1), 3.00, 2.60 (S2), 1.34, 2.51 (S3), 3.18, 2.11 (S4)]. Roex (p,w,t) filter shapes were fitted to the data as described by Yasin and Plack (2003) and represent the data reasonably well (see Table I for rms error values). The filter parameters were allowed to vary independently for the upper and lower skirts of the filter. In order to fit the roex functions, two additional free parameters were added ("BF" and L_{BF}). Parameter estimates were obtained using a least-squares minimization procedure using the MATLAB function "fminsearch." The equations which described the output of the roex filter for the lower and upper skirts were

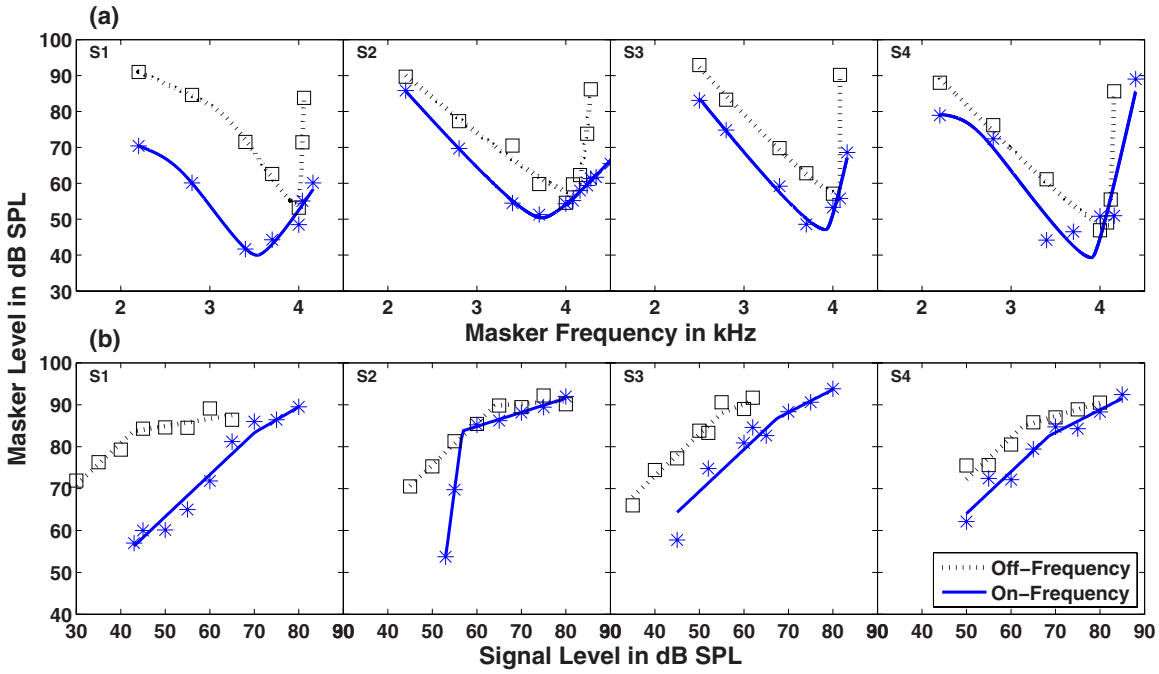


FIG. 3. (Color online) PTC (a) and GOM (b) results. The data are presented for each subject individually (S1–S4). Squares and asterisks represent the off-frequency and on-frequency precursor data, respectively. Dotted and solid lines are model fits to the data (see text).

$$W(f) = L_{\mathrm{BF}} - 10\log\left[(1-\mathrm{wl})\left(1+\mathrm{pl}\left|\frac{f-\mathrm{BF}}{\mathrm{BF}}\right|\right)e^{-\mathrm{pl}|f-\mathrm{BF}/\mathrm{BF}|} + \mathrm{wl}\left(1+\mathrm{tl}\left|\frac{f-\mathrm{BF}}{\mathrm{BF}}\right|\right)e^{-\mathrm{tl}|f-\mathrm{BF}/\mathrm{BF}|}\right], \quad (1)$$

$$W(f) = L_{\mathrm{BF}} - 10\log\left(1+\mathrm{pu}\left|\frac{f-\mathrm{BF}}{\mathrm{BF}}\right|\right), \quad (2)$$

where BF is the filter's best frequency, L_{BF} is the estimated output at BF, p ("pl" or "pu") and tl determine the filter slopes for the tip and tail, respectively, and wl delineates the intersection of pl and tl.

Table I presents the parameter estimates of the fitting procedure. The cells marked "NA" denote when a parameter did not appreciably improve the fit. Filter sharpness (Q_{10}) was estimated by dividing BF by the filter bandwidth at 10 dB above L_{BF}.

Three general patterns emerged from the PTC data. First, masker levels at threshold were lower in the on-frequency precursor condition at all masker frequencies (except for S4 at 4 kHz). Second, the estimated Q_{10} was reduced in all subjects in the on-frequency precursor condition; however, this effect was not significant ($t[3]=1.57$, p(one-tailed)$=0.106$). Finally, the estimated BF was consistently lower in the on-frequency precursor condition ($t[3]=4.32$, p(one-tailed)$=0.011$).

B. GOM data

The GOM results are plotted on the bottom row of Fig. 3. Masker level is plotted as a function of signal level. Error bars were omitted for clarity [the average standard deviations for the off and on-frequency data were 0.98, 2.57 (S1), 2.27, 3.13 (S2), 1.11, 0.81 (S3), 0.65, 1.64 (S4)]. Two lines were simultaneously fitted to the GOM data for each condition using a minimization procedure with three free parameters (G, c, and "BP1"). This procedure was modified from Yasin and Plack (2003). The first line was fitted to the lower leg of the data and was constrained to have a slope of 1. The y-intercept of this line was a free parameter (G) and can be interpreted as an estimate of gain. The second line was fitted to the upper leg of the data and its slope (c) can be interpreted as an estimate of compression. The lower breakpoint (BP1) delineated the end of the first line and the beginning of the second line. An additional breakpoint ("BP2") was initially a free parameter; however, it did not improve the fitting procedure so its value was fixed at 100 dB. The two-line model fits were given by

$$\mathrm{BM}(L_{\mathrm{in}}) = L_{\mathrm{in}} + G, \quad L_{\mathrm{in}} \le \mathrm{BP1}, \quad (3)$$

$$\mathrm{BM}(L_{\mathrm{in}}) = \mathrm{IBP1} + c(L_{\mathrm{in}} - \mathrm{BP1}), \quad \mathrm{BP1} < L_{\mathrm{in}} \le \mathrm{BP2}, \quad (4)$$

$$\mathrm{BM}(L_{\mathrm{in}}) = \mathrm{IBP2} + (L_{\mathrm{in}} - \mathrm{BP2}), \quad L_{\mathrm{in}} > \mathrm{BP2}, \quad (5)$$

where L_{in} is the input signal level, $\mathrm{BM}(L_{\mathrm{in}})$ is the estimated masker level at threshold, and the "internal breakpoints" are IBP1$=G+$BP1 and IBP2$=$IBP1$+c$(BP2$-$BP1). For one subject (S2), the lowest data point resulted in a poor model fit and an incorrect characterization of the breakpoint. The poor fit was largely due to the steep slope of the lower leg of the data. A similar steep slope was also observed in some subjects by Strickland (2008) and Krull and Strickland (2008) and may be related to the absolute threshold effects discussed by Plack and Skeels (2007). For this subject, the slope of the line fit to the lower leg of the data was allowed to vary. As a result of this minor change in the model, S2's data were not included in the statistical tests for gain.

The parameter estimates of the GOM fitting procedure are displayed in Table I. The on-frequency precursor condition produced lower estimates of gain ($t[2]=2.95$, p(one-tailed)$=0.049$) and higher estimates of the compression slope ($t[3]=-3.97$, p(one-tailed)$=0.028$) compared to the off-frequency precursor condition. There was no significant effect with respect to the breakpoint ($t[3]=-2.74$, p(one-tailed)$=0.138$). This was unexpected considering a previous study by Krull and Strickland (2008) that reported an increase in breakpoint under similar conditions. Three of the four subjects in the current study exhibited a fairly large increase in breakpoint, while the other (S2) showed a decrease. This reversal in the direction of the breakpoint for S2 may explain why this effect was not significant.

The estimates of compression slope are similar to previous studies with respect to the off-frequency precursor data. For example, Rosengard *et al.* (2005) reported values between 0.20 and 0.37 compared to a range of 0.07–0.46 in the current study. The slightly elevated values of compression slope in subjects 3 and 4 are likely a result of insufficient data above BP1 to obtain an accurate estimate. This is also true for the compression estimates in the on-frequency precursor condition for subjects 1 and 3.

Oxenham and Plack (1997) reported that the off-frequency GOM function became "more linear" between 40–50 dB, which can be interpreted as an estimate of BP1. The present study's estimates are somewhat higher than this range for three out of four of the subjects; however, similar high breakpoints have been observed in individual subjects in previous studies. For example, two subjects in Moore *et al.* (1999) (i.e., subjects JP and AG in Fig. 4) appear to have breakpoints between 55 and 60 dB SPL. Similarly, subject 2 in Oxenham and Plack (2000) showed no definable breakpoint up to the maximum signal threshold obtained (approximately 55 dB SPL). Finally, consistent with the current study, Krull and Strickland (2008) reported breakpoints above 60 dB SPL for two of their three subjects.[1]

Comparing the current data across experiments reveals a pattern that is consistent with a decrease in cochlear-amplifier gain in the on-frequency precursor condition. This pattern is characterized by lower estimates of gain, frequency selectivity, and BF. A similar pattern was noted by Strickland (2001), who measured frequency selectivity in simultaneous masking with and without a broadband precursor. The author interpreted these data as a reduction in cochlear gain and supported this interpretation using a simple BM I/O function model. Although frequency selectivity was measured, Strickland (2001) would not have been able to observe a decrease in BF due to the measurement method (i.e., notched-noise

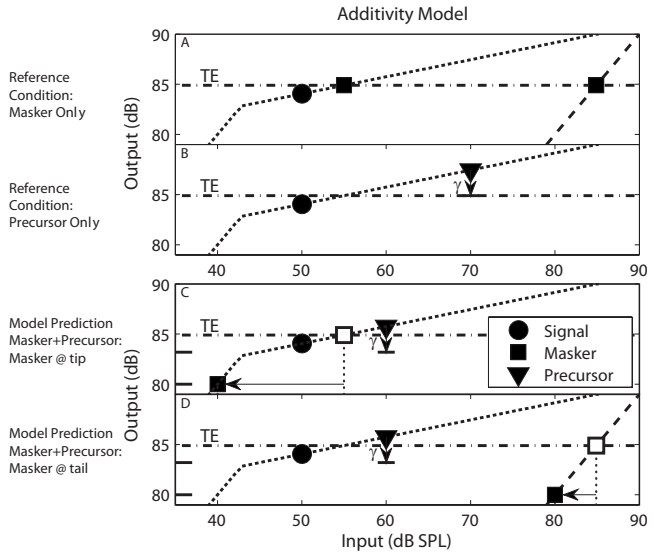


FIG. 4. Schematic representation of the additivity model using estimated BM I/O functions. The dotted line represents the I/O function fit to the GOM data in the off-frequency precursor condition [$BM_{unadapted}$(in)], while dashed lines represent a linear I/O function for the off-frequency masker. Panel (a) displays threshold for the signal (circle) in the presence of the masker (squares) alone. The data from this condition determined the signal-to-masker ratio at threshold. Panel (b) displays the threshold for the signal in the presence of the precursor (triangle) alone at L_{PA}. The constant γ represents the decay of the precursor energy during the silent interval between the signal and precursor. The value of γ was determined in panel (b) and assumed to be the same in panels (c) and (d). Panels (c) and (d) display how the precursor (at L_{PE}) and masker outputs (solid dashes on the y-axis) might add to produce an additivity-of-masking effect for the tip and tail frequencies. The horizontal line labeled "TE" represents the total energy needed to mask the signal (i.e., the addition of the masker and precursor outputs equal TE). The leftward pointing arrows represent the decrease in masker input needed to maintain a constant signal-to-masker+precursor ratio at the output of the BM I/O function. The open squares [and filled triangles in (c) and (d)] are the input(s) to the model and the closed squares are the predicted output.

method). In a recent forward masking study, Krull and Strickland (2008) estimated the BM I/O function using the off-frequency GOM technique. They reported a 5–20 dB reduction in gain across subjects when a 60 dB SPL precursor was present. This range is similar to Strickland (2008) and the current study, where a similar precursor produced 9–27 dB of gain reduction. It is noteworthy that physiological studies on the MOCR have observed reductions in BM gain of the same magnitude (Murugasu and Russell, 1996).

IV. GENERAL DISCUSSION AND MODELING

Although the pattern of results discussed above is consistent with a decrease in gain in the on-frequency precursor condition, there are other potential interpretations, the most prominent of which involves the additivity of masking (Oxenham and Moore, 1994; Plack and O'Hanlon, 2003; Plack *et al.*, 2006, 2007). Under this interpretation, the precursor and masker are viewed as two maskers whose intensities add after being processed by the auditory periphery. A typical additivity-of-masking experiment consists of two phases. During the first phase, the signal level is fixed and thresholds are measured for each masker individually. In the second phase, the maskers are presented together at the levels measured in the first phase, and the signal level at threshold is then measured. In a linear system, this approach should lead to a 3 dB increase in signal level relative to the first phase; however, threshold shifts are often much larger. This phenomenon is referred to as "excess masking" and has been interpreted as evidence of compression (Penner *et al.*, 1980). Under the additivity-of-masking assumption, the magnitude of excess masking is related to the amount of compression and the relative input levels of the two maskers for a given signal level. In experiment 2, the precursor level was fixed; therefore, as the signal level increases, any partial (or additive) masking from the precursor decreases. This argument suggests that excess masking should be largest at low signal levels where the partial masking from the precursor is highest. Qualitatively, this appears to be the case in all four subjects.

A modeling approach was used to evaluate the additivity-of-masking hypothesis. The experimental design in the current study differs from typical additivity-of-masking experiments; therefore, the model (hereafter referred to as the "additivity model") also differs and involves a few more assumptions. To compare the hypotheses of additivity and gain reduction, a "gain-reduction" model was tested as well. This model was also evaluated by predicting data from a previous study that was interpreted in terms of additivity of masking (Plack *et al.*, 2006).

For the current data set, the purpose of modeling was to predict the on-frequency precursor data (solid lines Fig. 3) by using the off-frequency precursor data (dotted lines Fig. 3) as a reference. For the PTC predictions, two masker frequencies were modeled, one at the signal frequency (4000 Hz) and the other well below the signal frequency (≈2200 Hz). These frequencies will be referred to as the "tip" and "tail" frequencies, respectively. They were chosen because their BM I/O functions were either measured in experiment 2 (tip frequency) or reasonably assumed to be linear with a slope of 1 (tail frequency). For the GOM predictions, only the additivity model was evaluated. Both additivity and gain-reduction models assumed no effect from the *off-frequency* precursor; therefore, in subsequent sections the term precursor will be reserved for the *on-frequency* precursor, unless otherwise stated.

A. Additivity model

Under the additivity-of-masking assumption, the effective intensities of the masker and precursor add internally at the output of the auditory periphery. Two thresholds were used as reference conditions and are shown schematically in Figs. 4(a) and 4(b).[2] These thresholds describe the individual internal representations of the precursor and masker when presented individually with the signal. Specifically, the precursor reference condition [Fig. 4(b)] was the threshold for the signal in the presence of the precursor at a 20 ms precursor-signal interval and is represented by the column L_{PA} in Table I. The masker reference condition [Fig. 4(a)] was the threshold measured for the off-frequency PTC at the tip and tail frequencies. All reference thresholds were assumed to be a function of the BM I/O function and a constant signal-to-masker ratio (or signal-to-precursor ratio for

the case of the precursor alone) at the output of the auditory periphery. An estimate of this ratio was obtained from the reference threshold for the masker alone and can be interpreted graphically as the vertical distance between the filled circle and filled squares in Fig. 4(a). For each listener, the fitted GOM function in experiment 2 (off-frequency precursor condition) served as the estimate of the BM I/O function [$BM_{unadapted}$(in)]. The internal representation of the precursor when presented alone with the signal ($IRP_{reference}$) was estimated by

$$IRP_{reference} = BM_{unadapted}(L_{PA}) + \gamma \quad (6)$$

in decibels, where L_{PA} is level of the precursor needed to mask the signal in the absence of the masker and γ is a constant representing the decrease in precursor output during the 20 ms silent interval between the precursor and signal. For the combined masker-precursor case, it was assumed that the signal was detected at a constant signal-to-masker +precursor ratio after peripheral processing; therefore, the total energy (TE) needed to mask the signal at threshold was described by

$$TE = 10\log(10^{IRP_{model}/10} + 10^{IRM_{model}/10}), \quad (7)$$

where IRP_{model} is Eq. (6) evaluated at L_{PE} and IRM_{model} is the predicted internal representation of the masker in the on-frequency precursor condition. IRM_{model} was solved from Eq. (7):

$$IRM_{model} = 10\log(10^{TE/10} - 10^{IRP_{model}/10}) \quad (8)$$

and then passed into the inverse of the BM I/O function to predict the masker threshold for the on-frequency precursor condition.

B. Gain-reduction model

In contrast to the additivity model, the gain-reduction model (Fig. 5) did not assume energetic masking from the precursor. Rather, it assumed that the precursor reduced the gain of the BM I/O function, after which the signal was detected at a constant signal-to-masker ratio. For each listener, the fitted GOM function in experiment 2 (on-frequency precursor condition) served as an estimate of the BM I/O function [$BM_{adapted}$(in)]. Note that for this model, the signal-to-masker ratio was determined solely by the masker (i.e., the precursor intensity did not add to the masker intensity before detection).

C. Model predictions and discussion

The predictions of the additivity and gain-reduction models are presented in Fig. 6. The PTC (top row) and GOM (bottom row) data were replotted. Dashed and solid lines represent the reference data (i.e., the input to the models) and the data to be predicted, respectively. Accurate model predictions lie near the solid line for a given listener. First consider the PTC predictions. The triangles and circles represent the additivity and gain-reduction model predictions, respectively (rms error: additivity model=8.17 dB; gain-reduction model=3.40 dB). As expected both models predicted a reduction in masker level at threshold relative to the reference

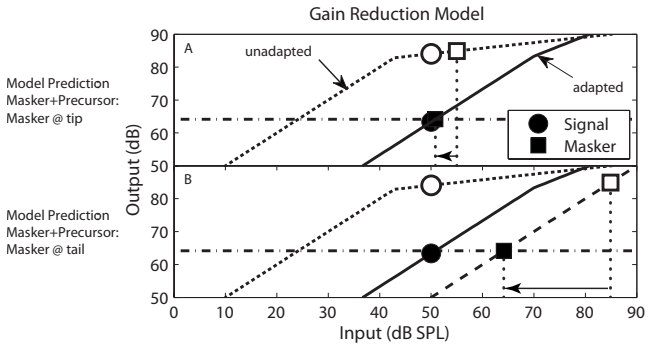


FIG. 5. Schematic representation of the gain-reduction model. This model assumed the on-frequency precursor reduced BM gain (or equivalently "adapted" the on-frequency I/O function). Two masker frequencies are schematized, one at the same frequency as the signal (a) and the other far below the signal frequency (b). Within each panel, the masker (squares) and signal (circles) are plotted for the masker alone condition (open symbols), or for the masker+precursor condition (closed symbols). Notice that the masker alone condition assumes that the signal and on-frequency masker operate on the unadapted I/O function [$BM_{unadapted}$(in), dotted lines], while the masker+precursor condition assumes that the signal and on-frequency masker operate on the adapted I/O function [$BM_{adapted}$(in), solid lines]. The dashed line in (b) represents the linear I/O function for the off-frequency masker. The horizontal line represents the masker output needed to maintain a constant signal-to-masker ratio at the output of the adapted I/O function.

data. At the tip frequency, the models are nearly equally accurate (except for S1, where the additivity model is poor). In contrast, a deviation between models is observed at the tail frequency, where the gain-reduction model is more accurate.[3] Krull and Strickland (2008) reported a similar finding when attempting to model precursor effects on off-frequency GOM under the additivity assumption. Their additivity model underestimated the effect of the precursor in all three of their subjects. Consistent with their findings, the current additivity simulations involving off-frequency masking (i.e., the tail frequency condition of the PTC experiment and the entire GOM experiment) underestimated the effect of the precursor. These results suggest that although additivity of masking explains data for two consecutive forward maskers with energy at the signal frequency, it may not explain data for two consecutive forward maskers having energy at and well below the signal frequency [however, see Arifianto and Plack (2008)].

It is important to note that the gain-reduction model does not to attempt to predict the shift in BF observed in the data. In order to predict this shift, the model would need an estimate of the BM I/O function at each masker frequency. Given the current data, such estimates are not available; however, a decrease in BF is often associated with an excitation pattern dominated by the passive mechanics of the BM (Ruggero *et al.*, 1997). This domination may occur as a result of high input levels (Moore *et al.*, 2002) or due to a decrease in the active process. For example, Yasin and Plack (2003) observed a decrease in BF in their PTC data when a suppressor was present, which is consistent with suppression functioning as an instantaneous decrease in the active process. The decrease in BF observed in the current experiment may also reflect a decrease in the active process given the low signal level used to measure the PTCs.

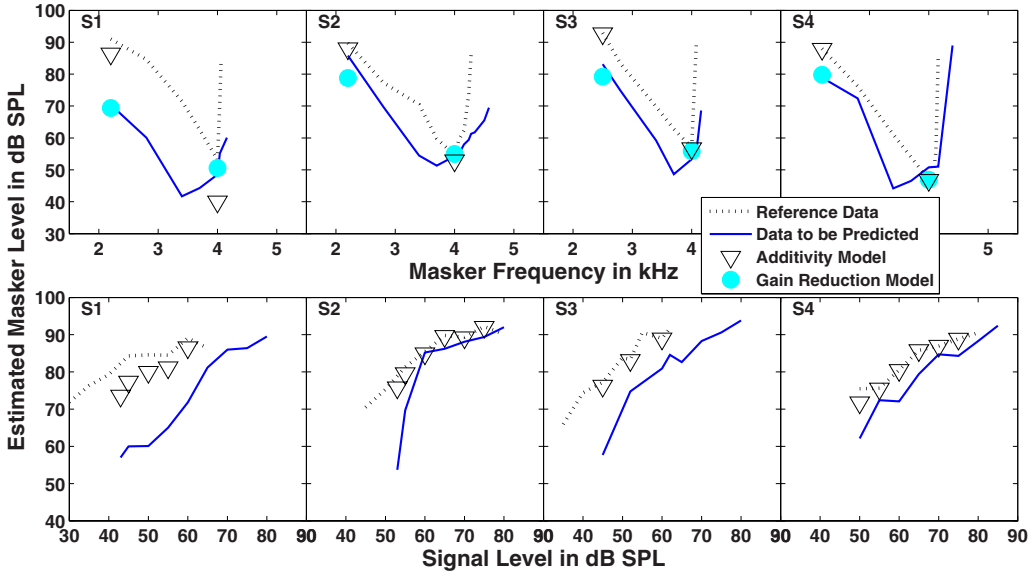


FIG. 6. (Color online) Model prediction for the PTC (top row) and GOM (bottom row) experiments. The predictions of the additivity (PTC and GOM data) and the gain-reduction (PTC data only) models are displayed as open triangles and filled circles, respectively. The models were intended to predict the on-frequency precursor data (asterisks in Fig. 3), which are replotted as a solid line. The dashed lines are the off-frequency data replotted and serve as a reference.

D. Modeling previous data

To further test the gain-reduction model, predictions were obtained for the mean data in Plack *et al.* (2006). This study involved measuring signal threshold in the presence of one or two temporally non-overlapping on-frequency forward maskers (M1 and M2). Using an additivity-of-masking model, Plack *et al.* (2006) were able to predict signal thresholds for the combined masker condition (M1+M2), by referencing signal thresholds from the separate masker conditions (M1 or M2 alone) and assuming a constant signal-to-masker ratio. A similar approach was taken in the current study; however, it was assumed that M1 reduced the gain of the I/O function rather than adding with M2 at the output of the BM. Figure 7 displays the predictions of the gain-reduction model and the I/O function which minimized the error of these predictions. The equation for this I/O function was similar to that used in Plack *et al.* (2006), except in units of decibels and with the additional constant d:

$$f(x) = ax^3 + bx^2 + cx + d, \tag{9}$$

where x is the input signal level (dB SPL) and a, b, and c are coefficients. The assumptions of the model were as follows: (1) The signal was detected at a constant signal-to-masker ratio at the output of the auditory periphery, (2) signal thresholds for M2 alone were a result of an I/O function with high gain (dotted line in Fig. 7), and (3) signal thresholds for M1+M2 were a result of an I/O function with relatively lower gain (solid line in Fig. 7). The coefficients for the "high-gain" I/O function were fixed at the values measured for the mean data in Plack *et al.* (2006) ($a=4.3\times10^{-5}$, $b=-9.9\times10^{-3}$, $c=0.913$, and $d=0$). Gain reduction was simulated by allowing the parameter d to vary, while constraining

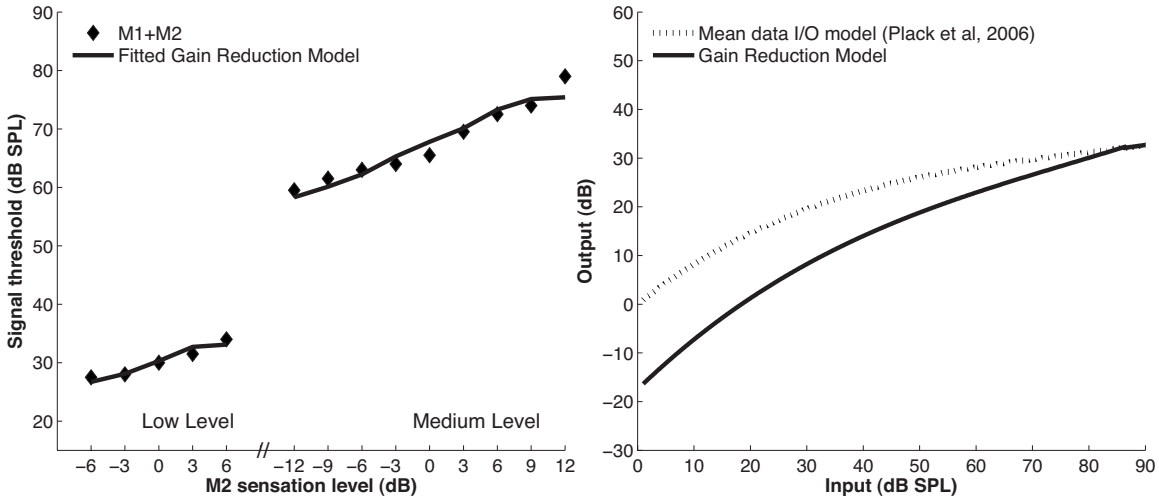


FIG. 7. Gain-reduction model predictions of Plack *et al.* (2006) (mean data) and the I/O function which minimized the error of the predictions. Parameter estimates of the gain-reduction model were $d=-17.48$, $x_{eq}=85.47$, and $k=-9.81$

the output to be equal to the high-gain model above a certain input level (x_{eq}). This constraint produced the following relationship for the coefficient c:

$$c = [f(x_{eq}) - ax_{eq}^3 - bx_{eq}^2 - d]/x_{eq}, \quad (10)$$

where $f(x_{eq})$ is the output of the high-gain polynomial model at x_{eq}. The values of a and b were constrained to be equal to those of the high-gain model; therefore, the three free parameters in the gain-reduction model were d, x_{eq}, and the threshold signal-to-masker ratio (k). As shown in Fig. 7, the gain-reduction model captures the data reasonably well. These results support the previous observation that additivity and gain-reduction models are equally accurate when the maskers (or the masker and precursor) have the same frequency.[4]

V. SUMMARY AND CONCLUSIONS

The experiments and model simulations presented in this study support the hypothesis that a reduction in gain can account for the effects of a precursor (or second masker) in the current study and at least one previous study. This conclusion is consistent with previous studies in forward and simultaneous masking which have also modeled masking data as a reduction in gain (Strickland, 2001; Oxenham, 2001; Strickland, 2004; Strickland and Krishnan, 2005; Strickland, 2008; Krull and Strickland, 2008). This study also showed that additivity of masking can account for some of the data in the experiments; however, for much of the data the size of the precursor effect was larger than predicted by additivity.

An interesting dichotomy emerges when comparing simultaneous and forward masking in terms of precursor effects. In simultaneous masking, a precursor usually improves detectability relative to thresholds without a precursor, while the opposite is true in forward masking. This dichotomy suggests that the human auditory system is designed for optimal performance in simultaneous masking tasks. Such a design seems appropriate given that most "everyday" listening tasks involve detecting and understanding signals embedded in background noise. In such environments, the MOCR may adjust gain to reduce the influence of competing signals and their effects on neural adaptation and transmitter depletion (Guinan, 2006). Similarly, such an adjustment may help to shift the rate-level functions of auditory neurons to optimize encoding of the signal of interest (Dean *et al.*, 2005).

ACKNOWLEDGMENTS

We would like to thank the associate editor Magdalena Wojtczak and reviewers Chris Plack and Andrew Oxenham for their helpful comments and suggestions on previous versions of this manuscript. Similarly, we would like to thank Vidya Ganesh for her comments on the revised manuscript. This work was funded in part by NIH (NIDCD) Grant Nos. RO1-DC008327 and T32-DC00030.

[1]The frequency response of the ER-2 headphones used in the current study may contribute to the high breakpoints. These headphones produce a flat frequency response at the level of the eardrum. Many of the previous studies have used circumaural headphones such as the Sennheiser HD580 (Rosengard *et al.*, 2005), which may produce a slightly higher output in the 2–6 kHz region (the frequency region of our signal) when calibrated in a similar manner.

[2]This schematic is based on the data from S1 with a slight modification to add visual clarity. This modification involved slightly increasing the masker level at threshold, as well as slightly decreasing the value of L_{PA}.

[3]It is not surprising that the gain-reduction model predicts the data at the tail frequency since the estimated I/O functions used in the model were derived under similar stimuli conditions as the data being predicted. These data could have been equally well predicted by a BM model where gain varied as a parameter; however, to limit the number of the parameters to the model, the GOM functions were used instead.

[4]The model simulations were limited to the M1+M2 data. Although the model could be revised to include M1 predictions, such a revision would require adding additional assumptions and parameters. In an effort to reduce the model's parameters and assumptions, we limited our predictions to the M1+M2 data.

Arifianto, D., and Plack, C. J. (**2008**). "Behavioral evidence for off-frequency compression at 4 kHz," J. Acoust. Soc. Am. **123**, 3858.
Backus, B. C., and Guinan, J. J. (**2006**). "Time-course of the human medial olivocochlear reflex," J. Acoust. Soc. Am. **119**, 2889–2904.
Bacon, S. P., and Jesteadt, W. (**1987**). "Effects of pure-tone forward masker duration on psychophysical measures of frequency selectivity," J. Acoust. Soc. Am. **82**, 1925–1932.
Bacon, S. P., and Moore, B. C. (**1986**). "Temporal effects in masking and their influence on psychophysical tuning curves," J. Acoust. Soc. Am. **80**, 1638–1645.
Bacon, S. P., Repovsch-Duffey, J. L., and Liu, L. (**2002**). "Effects of signal delay on auditory filter shapes derived from psychophysical tuning curves and notched-noise data obtained in simultaneous masking," J. Acoust. Soc. Am. **112**, 227–237.
Bacon, S. P., and Viemeister, N. F. (**1985**). "The temporal course of simultaneous tone-on-tone masking," J. Acoust. Soc. Am. **78**, 1231–1235.
Cooper, N. P., and Guinan, J. J. (**2006**). "Efferent-mediated control of basilar membrane motion," J. Physiol. (London) **576**, 49–54.
Dean, I., Harper, N. S., and McAlpine, D. (**2005**). "Neural population coding of sound level adapts to stimulus statistics," Nat. Neurosci. **8**, 1684–1689.
Guinan, J. J. (**2006**). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," Ear Hear. **27**, 589–607.
Guinan, J. J., Jr., and Gifford, M. L. (**1988**). "Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. III. Tuning curves and thresholds at CF," Hear. Res. **37**, 29–45.
Hicks, M. L., and Bacon, S. P. (**1992**). "Factors influencing temporal effects with notched-noise maskers," Hear. Res. **64**, 123–132.
Kidd, G., Mason, C. R., and Feth, L. L. (**1984**). "Temporal integration of forward masking in listeners having sensorineural hearing loss," J. Acoust. Soc. Am. **75**, 937–944.
Kimberley, B. P., Nelson, D. A., and Bacon, S. P. (**1989**). "Temporal overshoot in simultaneous-masked psychophysical tuning curves from normal and hearing-impaired listeners," J. Acoust. Soc. Am. **85**, 1660–1665.
Krull, V., and Strickland, E. A. (**2008**). "The effect of a precursor on growth of forward masking," J. Acoust. Soc. Am. **123**, 4352–4357.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Moore, B. C., Alcantara, J. I., and Glasberg, B. R. (**2002**). "Behavioural measurement of level-dependent shifts in the vibration pattern on the basilar membrane," Hear. Res. **163**, 101–110.
Moore, B. C., 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.
Moore, B. C., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). "Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism," J. Acoust. Soc. Am. **106**, 2761–2778.
Murugasu, E., and Russell, I. J. (**1996**). "The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea," J. Neurosci. **16**, 325–332.
Oxenham, A. J. (**2001**). "Forward masking: Adaptation or integration?" J. Acoust. Soc. Am. **109**, 732–741.
Oxenham, A. J., and Bacon, S. P. (**2004**). in *Compression: From Cochlea to Cochlear Implants*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper

(Springer, New York), pp. 62–106.
Oxenham, A. J., and Moore, B. C. (**1994**). "Modeling the additivity of nonsimultaneous masking," Hear. Res. **80**, 105–118.
Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.
Oxenham, A. J., and Plack, C. J. (**2000**). "Effects of masker frequency and duration in forward masking: Further evidence for the influence of peripheral nonlinearity," Hear. Res. **150**, 258–266.
Penner, M. J., Shiffrin, R. M., and Shiffrin, R. M. (**1980**). "Nonlinearities in the coding of intensity within the context of a temporal summation model," J. Acoust. Soc. Am. **67**, 617–627.
Plack, C. J., Carcagno, S., and Oxenham, A. J. (**2007**). "A further test of the linearity of temporal summation in forward masking," J. Acoust. Soc. Am. **122**, 1880–1883.
Plack, C. J., and O'Hanlon, C. G. (**2003**). "Forward masking additivity and auditory compression at low and high frequencies," J. Assoc. Res. Otolaryngol. **4**, 405–415.
Plack, C. J., Oxenham, A. J., and Drga, V. (**2006**). "Masking by inaudible sounds and the linearity of temporal summation," J. Neurosci. **26**, 8767–8773.
Plack, C. J., and Skeels, V. (**2007**). "Temporal integration and compression near absolute threshold in normal and impaired ears," J. Acoust. Soc. Am. **122**, 2236–2244.
Rosengard, P. S., Oxenham, A. J., and Braida, L. D. (**2005**). "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **117**, 3028–3041.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.
Schmidt, S., and Zwicker, E. (**1991**). "The effect of masker spectral asymmetry on overshoot in simultaneous masking," J. Acoust. Soc. Am. **89**, 1324–1330.
Strickland, E. A. (**2001**). "The relationship between frequency selectivity and overshoot," J. Acoust. Soc. Am. **109**, 2062–2073.
Strickland, E. A. (**2004**). "The temporal effect with notched-noise maskers: Analysis in terms of input-output functions," J. Acoust. Soc. Am. **115**, 2234–2245.
Strickland, E. A. (**2008**). "The relationship between precursor level and the temporal effect," J. Acoust. Soc. Am. **123**, 946–954.
Strickland, E. A., and Krishnan, L. A. (**2005**). "The temporal effect in listeners with mild to moderate cochlear hearing impairment," J. Acoust. Soc. Am. **118**, 3211–3217.
von Klitzing, R., and Kohlrausch, A. (**1994**). "Effect of masker level on overshoot in running- and frozen-noise maskers," J. Acoust. Soc. Am. **95**, 2192–2201.
Yasin, I., and Plack, C. J. (**2003**). "The effects of a high-frequency suppressor on tuning curves and derived basilar-membrane response functions," J. Acoust. Soc. Am. **114**, 322–332.
Zwicker, E. (**1965**). "Temporal effects in simultaneous masking and loudness," J. Acoust. Soc. Am. **38**, 132–141.

Auditory stream formation affects comodulation masking release retroactively[a)]

Torsten Dau[b)]
Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark

Stephan Ewert
Medizinische Physik, Institut für Physik, Universität Oldenburg, D-26111 Oldenburg, Germany

Andrew J. Oxenham
Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

(Received 31 October 2007; revised 23 January 2009; accepted 26 January 2009)

Many sounds in the environment have temporal envelope fluctuations that are correlated in different frequency regions. Comodulation masking release (CMR) illustrates how such coherent fluctuations can improve signal detection. This study assesses how perceptual grouping mechanisms affect CMR. Detection thresholds for a 1-kHz sinusoidal signal were measured in the presence of a narrowband (20-Hz-wide) on-frequency masker with or without four comodulated or independent flanking bands that were spaced apart by either 1/6 (narrow spacing) or 1 octave (wide spacing). As expected, CMR was observed for the narrow and wide comodulated flankers. However, in the wide (but not narrow) condition, this CMR was eliminated by adding a series of gated flanking bands after the signal. Control experiments showed that this effect was not due to long-term adaptation or general distraction. The results are interpreted in terms of the sequence of "postcursor" flanking bands forming a perceptual stream with the original flanking bands, resulting in perceptual segregation of the flanking bands from the masker. The results are consistent with the idea that modulation analysis occurs within, not across, auditory objects, and that across-frequency CMR only occurs if the on-frequency and flanking bands fall within the same auditory object or stream.

[DOI: 10.1121/1.3082121]

PACS number(s): 43.66.Dc, 43.66.Mk, 43.66.Ba [BCM] Pages: 2182–2188

I. INTRODUCTION

The audibility of a target sound embedded in another masking sound can be improved by adding sound energy that is remote in frequency from both the masker and the target (Hall *et al.*, 1984). This effect is known as comodulation masking release (CMR) and is observed when the remote sound and the masker share coherent patterns of amplitude modulation. Most ecologically relevant sounds, such as speech and animal vocalizations, have coherent amplitude modulation patterns across different frequency regions, suggesting that the detection and recognition advantages conveyed by such coherent modulations may play an important role in our ability to deal with natural complex acoustic environments (e.g., Klump, 1996; Nelken *et al.*, 1999).

CMR has been measured in two ways. The first, often referred to as the "band-widening experiment," is to use a single band of noise, centered around the signal frequency, as a masker and to compare thresholds for modulated and unmodulated noise maskers as a function of the masker bandwidth (e.g., Hall *et al.*, 1984; Carlyon *et al.*, 1989). For the random noise, with irregular fluctuations in amplitude that are independent of different frequency regions, the signal threshold increases as the masker bandwidth increases up to about the critical bandwidth at that frequency and then remains constant, in broad agreement with the classical power spectrum model of masking (Fletcher, 1940; Patterson and Moore, 1986). For the modulated noise, a random noise that is amplitude modulated using a lowpass filtered noise as a modulator, the pattern of results is quite different. Here, signal thresholds typically decrease as the bandwidth increases beyond about 100 Hz (for a signal frequency of 2 kHz); thus, increasing the masker energy and bandwidth makes the signal easier to detect. These findings suggest that listeners may compare the outputs of different auditory filters to enhance signal detection. The fact that the decrease in threshold with increasing bandwidth only occurs with the modulated noise indicates that fluctuations in the masker are critical and that the fluctuations need to be correlated across frequency bands.

The second method is to use a masker consisting of several narrow masker bands of noise, typically with bandwidths between 20 and 50 Hz, which have relatively slow inherent amplitude fluctuations. One band is centered at the signal frequency (on-frequency band) and one or more other bands (flanking bands) are spectrally separated from the on-frequency band (e.g., Hall *et al.* 1984; Schooneveldt and Moore, 1987). When the flanking bands are uncorrelated with the on-frequency band, there is sometimes a slight el-

[a)]Portions of these data were presented at the 2003 International Symposium on Hearing in Paris, France, the proceedings of which are published as Dau *et al.*, 2005.
[b)]Electronic mail: tda@elektro.dtu.dk

evation but typically no effect on signal threshold, so long as the flanking bands are not so close in frequency as to produce direct masking. However, when the amplitude fluctuations of the flanking bands are correlated with those of the on-frequency band, the addition of the flanking bands can produce a release from masking (Hall *et al.*, 1984; Schooneveldt and Moore, 1987; Cohen and Schubert, 1987). CMR has been found even if the signal and on-frequency band are presented to one ear and the flanking bands to the other ear (Schooneveldt and Moore, 1987; Cohen and Schubert, 1987; Buss and Hall, 2008).

Even though CMR has been investigated in many studies, the underlying mechanisms are still not clear. It was originally assumed that CMR results from across-channel comparisons of temporal envelopes (e.g., Buus, 1985). However, there is evidence that within-channel cues, i.e., information from only the one peripheral channel tuned to the signal frequency, can account for a considerable part of the effect in some conditions, suggesting that within-channel processing can lead to an overestimation of "true" across-channel CMR (Schooneveldt and Moore, 1987). This conclusion was supported by simulations of data from a bandwidening experiment, using a modulation filterbank analysis of the stimuli at the output of the auditory filter tuned to the signal frequency (Verhey *et al.*, 1999; Piechowiak *et al.*, 2007). Additionally, for the CMR experiments using flanking bands, McFadden (1986) pointed out that it is imprecise to assume that one channel is receiving only the on-frequency band plus signal and another channel is receiving only the flanking band. Often, the two bands will be incompletely resolved. When this happens, the resulting waveform may contain envelope fluctuations resulting from beats between the carrier frequencies of the on-frequency and the flanker bands. These beats can facilitate signal detection without across-channel comparisons being involved (Schooneveldt and Moore, 1987). Thus, at least part of the CMR in many situations can be explained in terms of the use of within-channel rather than across-channel cues.

The authors of several studies have suggested that higher-level processes, such as object formation, may be involved in CMR, because certain stimulus manipulations designed to perceptually segregate the masker from the flanking bands have resulted in a reduction or elimination of CMR (McFadden and Wright, 1992; Grose and Hall, 1993). However, when manipulating perceptual grouping, it is often difficult to rule out mechanisms such as neural inhibition, forward suppression, or adaptation (Calford and Semple, 1995; Ulanovsky *et al.*, 2004; Wehr and Zador, 2005) that might at least partly be based on more peripheral processing. For example, neuronal adaptation, the decline over time of neural responses during sensory stimulation, might have affected the neural representation of the flanking masker bands in the experimental conditions of Grose and Hall (1993). In their study, CMR was reduced or eliminated either by gating the flanking bands on earlier and gating off later than the on-frequency masker band, or by presenting a series of precursor bands at the same frequencies as the flanking masker bands to perceptually segregate the on-frequency from the flanking masker bands. In both cases, it is at least conceivable that the main effect of the precursors or leading onset asynchronies was to reduce the neural response to the flanking masker bands.

To exclude adaptation as a possible basis for the reduction or elimination of CMR, the present study focused on sounds that occurred *after* the target in time. Sounds occurring after a target and masker interval could in principle affect their perception by, for instance, binding with the flanking bands to form a separate perceptual stream (Dannenbring and Bregman, 1978; Bregman, 1990). Our hypothesis was that if across-frequency modulation analysis (and hence CMR) occurs primarily within auditory objects, then CMR could be eliminated by sounds that occur after the target, so long as these sounds are successfully segregating the on-frequency masker and the flanking masker bands into different auditory objects, thereby disrupting the across-frequency (but within-object) modulation processing. On the other hand, if across-frequency modulation processing is a lower-level or "automatic" process that is not governed by auditory grouping mechanisms, then sounds occurring after the target in time should not affect CMR.

II. RETROACTIVE STIMULUS EFFECTS ON CMR

A. Method

1. Listeners

Six normal-hearing listeners ranging in age from 25 to 39 years participated in the experiments. Two of the listeners were the first and second authors. All listeners received several hours of listening experience prior to the final data collection.

2. Apparatus and stimuli

Listeners were seated in a double-walled sound attenuating booth in front of a computer keyboard and monitor. The stimuli were presented diotically via Sennheiser HD580 headphones. Signal generation and presentation during the experiments were controlled by computer using the AFC software package for MATLAB, developed at the University of Oldenburg and the Technical University of Denmark. The stimuli were digitally generated at a sampling rate of 32 kHz and converted to analog signals by a high-quality 24-bit sound card (RME DIGI96/8 PAD).

Figure 1 shows schematic spectra of the stimuli: The target (a 1-kHz tone) was masked by a narrow (20-Hz wide) band of noise centered at 1 kHz. Four flanking bands of noise (each 20-Hz wide) were presented with temporal envelopes that were either random (condition R) and thus uncorrelated with that of the on-frequency masker, or coherent (condition C) with the masker envelope. The envelope fluctuations, or modulations, of the bands are indicated by the different shades of gray. The coherent across-frequency modulation of condition C was expected to enhance the audibility of the target, and hence reduce its detection threshold, relative to its threshold in condition R. The novel conditions investigated the retroactive influence of stimulus presentation on CMR. Several additional bursts of noise, termed postcursors, were presented at the frequencies of the original flanking bands (conditions PR and PC). The enve-

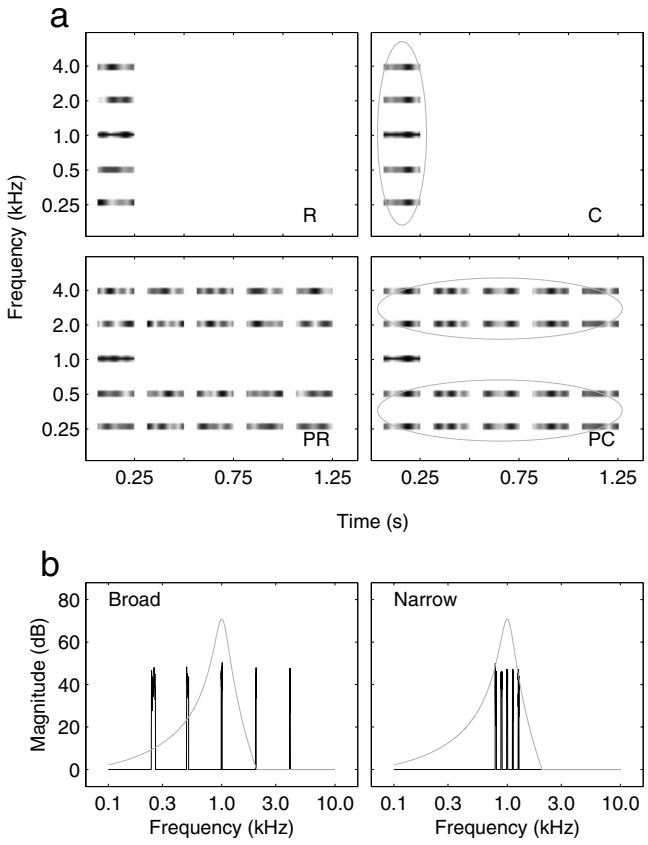


FIG. 1. Schematic representation of the experimental conditions. (a) Four conditions with a 1-kHz target tone (black horizontal line) masked by a noise band centered at 1-kHz with random flankers (R), comodulated flankers (C), random flankers followed by four postcursors (PR), and comodulated flankers followed by four comodulated postcursors (PC). The shades of gray indicate the distribution of envelope fluctuations in the masker and flanker bands. (b) Power spectra for the broadband configuration (with one-octave spacing between the noise bands) and narrowband configuration (with one-sixth octave spacing). The gray curves represent the magnitude response of the auditory filter centered at the target frequency.

lope coherence of the postcursors was the same as that for the original flanking masker bands. The postcursors were designed to "capture," and to form a single auditory stream with, the original flanking bands and thus to perceptually segregate them from the masker (Bregman and Pinker, 1978), as shown schematically in Fig. 1(a) (gray ellipses).

The experiment was performed using two spectral configurations [Fig. 1(b)]. In the broadband configuration, the noise bands were centered at 250, 500, 1000, 2000, and 4000 Hz, i.e., with one-octave spacing between the bands such that they primarily stimulated separate auditory filters along the tonotopic axis. The gray curve in Fig. 1(b) indicates the magnitude transfer function of a gammatone bandpass filter (e.g., Patterson *et al.*, 1995) tuned to the signal frequency (1 kHz). In the narrowband configuration, the noise bands were centered at 794, 891, 1000, 1123, and 1260 Hz, representing a sixth-octave spacing centered around the target frequency. In this case, within-channel processes were likely to play a strong role (Schooneveldt and Moore, 1987; Verhey *et al.*, 1999), because all the components fell within the same frequency region. To the extent that the effect of the postcursors is limited to across-frequency processing, they should not affect CMR for the narrowband configuration. Hence, the narrowband configuration acted as a control condition for any potential non-specific distraction or interference effects produced by the postcursors.

In both cases, the noise bands were generated in the time domain as independent Gaussian noise tokens for each of the presentation intervals. The noise tokens were restricted to the appropriate bandwidth in the spectral domain via a Fourier transform. Comodulated noises were frequency-shifted versions of the masker band at 1000 Hz. The level of each of the noise bands was 60 dB sound pressure level (SPL). The four postcursors at the flanking-band frequencies all had the same duration and level as the masker bands (187.5 ms, including 20-ms raised-cosine ramps) and were separated by gaps of 62.5 ms, giving an overall repetition period of 250 ms.

3. Procedure

An adaptive, three-interval, three-alternative forced-choice procedure was used in conjunction with a 1-up, 2-down tracking rule to estimate the 70.7% correct point on the psychometric function (Levitt, 1971). The intervals were marked on a computer monitor and feedback was provided after each trial. Listeners responded via the computer keyboard or mouse. The initial step size of the target level was 8 dB, which was reduced to 4 and 2 dB after the second and fourth reversals, respectively. The adaptive run then continued for a further six reversals at the final step size, and threshold was defined as the mean of the levels at those last six reversals. Four threshold estimates were obtained and averaged from each listener in each condition. The intra-individual standard deviations were typically around 1–2 dB and rarely exceeded 4 dB. Final thresholds reported here are the mean across listeners, who all showed comparable patterns of results across conditions. Typical individual differences were around 2–3 dB and maximally reached 5–6 dB.

B. Results and discussion

The experimental data are shown in Fig. 2. In the broadband configuration (circles and filled bars), the results for conditions R and C were as expected from previous studies (Schooneveldt and Moore, 1987): The target threshold in the presence of the masker was significantly lower for the coherently modulated flanking bands than for the randomly modulated flanking bands [paired t-test; $t(5)=7.21$, $p<0.001$]. This difference in threshold, reflecting the amount of CMR, was 6.1 dB. However, this CMR was eliminated when the postcursors were added: Thresholds in condition PC were not significantly different from those in condition R [paired t-test; $t(5)=0.73$, $p=0.50$]. Similarly, there was no significant difference between thresholds in conditions PR and PC [paired t-test; $t(5)=0.59$, $p=0.58$], confirming the lack of effect of coherent amplitude modulations when the postcursors were present. The elimination of the CMR by sounds occurring after the target suggests that the postcursors led to perceptual segregation of the flanking bands from the masker, so that the coherent modulations in the flanking bands were no longer processed with those of the masker.

The results from the narrowband configuration (Fig. 2, squares and open bars) show that the effect of the postcursors

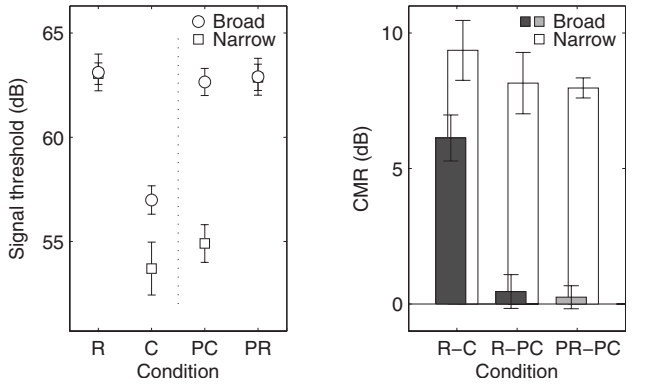


FIG. 2. Mean masked thresholds for the target tone (left) and amount of CMR (right) for the broadband (circles and filled bars) and the narrowband configurations (squares and open bars). Error bars denote one standard error across subjects. Conditions are indicated on the abscissa (R=random modulations of the flanking bands, C=comodulated flanking bands, PC =postcursors with comodulated flanking bands, and PR=postcursors with randomly modulated flanking bands). The amount of CMR, defined as the difference between thresholds in the random and the comodulated conditions, is indicated for the standard condition without postcursors, R-C, and for the conditions with postcursors (R-PC and PR–PC).

is unlikely to be a non-specific distraction effect: Here, the average target threshold in the presence of the masker (R) was 9.4 dB lower for the coherently modulated than for the randomly modulated flanking bands [paired t-test; $t(5)=8.46$, $p<0.001$], as expected (Schooneveldt and Moore, 1987). However, in contrast to the broadband configuration, no significant reduction in CMR was produced when the postcursors were added [C-PC, $t(5)=-1.06$, $p=0.34$]. When compared to the random condition with postcursors, PR-PC (right open bar), the amount of CMR was 8.0 dB, in contrast to a non-significant 0.25 dB in the broadband configuration (right light-gray bar). Thus, postcursors eliminated CMR in the broadband configuration where CMR is likely to be based on across-frequency processing, but did not significantly affect target detection in the narrowband configuration, where CMR is more likely to be based on within-channel cues. In other words, the postcursors were successful at eliminating CMR only when CMR was likely to be based on a true across-frequency analysis of coherent modulations occurring in remote frequency bands.

Overall, the results are difficult to account for in terms of traditional neuronal adaptation or inhibition mechanisms because the critical sound components (the postcursors) occurred after the presentation of both the masker and target. However, it is possible in principle that the postcursors in a given trial affected the representation of the flankers in the following trial, via some form of long-term adaptation. In other words, despite the temporal gaps between successive trials, the postcursors may have influenced the response to the flanking bands in the next trial, which in turn may have reduced CMR. Long-term adaptation effects have been observed in the auditory pathways, particularly at higher levels, such as cortex (e.g., Ulanovsky *et al.*, 2004; Altmann *et al.*, 2007). Another possibility is that the postcursors induced some distraction effect that selectively impaired performance in the wideband, but not the narrowband conditions. For instance, if attention were exogenously diverted toward the frequency regions of the flanking bands by the postcursors, and if this shift in attention affected signal detection, then this would be expected to selectively affect results for the wideband condition. Such an attentional effect would be less likely to affect results for the narrowband condition, because the frequency region of the flankers was close to that of the target. Both these possibilities were addressed in the following control conditions.

III. CONTROL CONDITIONS: MISSING AND OFF-FREQUENCY POSTCURSORS

A. Rationale

To address the possibility that the effect of the postcursors was due to longer-term effects on flankers in following trials, the first burst of postcursors following the flankers was replaced with a silent gap. The gap was expected to reduce or eliminate the perceptual grouping of the flankers with the postcursors, but would not be expected to eliminate any longer-term adaptation effects. To address the possibility that the effect of the postcursors was due to an attentional shift away from the target frequency, a second control condition was run in which the postcursors were shifted by a half-octave away from the frequencies of the flankers. In this case, attention would still be expected to shift from the target frequency, but the postcursors would no longer be expected to form a perceptual stream with the flanking masker bands. Thus, if the effect of the postcursors in experiment 1 was primarily due to an exogenous attentional shift, then CMR should be reduced even when the center frequencies of the postcursors are shifted; however, if the effect was due to perceptual grouping of the flanking maskers and postcursors into a single stream, then CMR should remain when the center frequencies are shifted, because the postcursors should no longer form a stream with the flanking masker bands.

B. Methods

The target and the flankers were the same as for the broadband configuration in the main experiment. The same listeners took part, and the apparatus and procedures for estimating thresholds were also the same. The left panel of Fig. 3(a) shows schematic spectrograms of the stimuli for the first control condition, where the first of the four postcursors was removed (gap-postcursor condition, GP); the right panel of Fig. 3(a) shows schematic spectrograms of the stimuli for the second control condition, where the center frequencies of the postcursors were shifted relative to the flanker center frequencies. The off-frequency postcursors were positioned with half-octave separation from the respective flanking bands, i.e., at 354, 707, 1414, and 2828 Hz.

C. Results and discussion

The data in Fig. 3(b) show that CMR is not affected by the postcursors, if the first postcursor is omitted. The threshold obtained in the comodulated condition with gap-postcursor (GPC) was not significantly different from the threshold obtained in the original condition C [from Fig. 2; $t(5)=-0.76$, $p=0.483$]. CMR was 5.5 dB when defined as

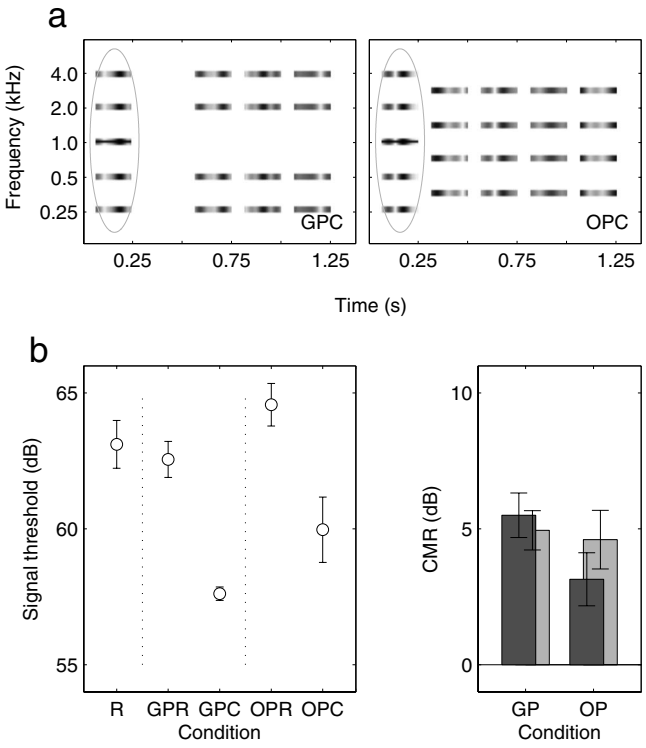


FIG. 3. Control stimulus conditions to further test the hypothesis that CMR is associated with auditory grouping. (a) Conditions with gap and postcursors (GPC, left) where the first postcursor was eliminated, and with off-frequency postcursors (OPC, right) where the postcursors were presented at intermediate frequencies. The last letter C in the abbreviations indicates that the comodulated condition is shown. (b) Mean masked thresholds and standard errors (left panel) are shown for the different stimulus conditions. The corresponding amounts of CMR for the conditions GP and OP are shown in the right panel. Dark-gray bars indicate the conditions R-GPC and R-OPC and light-gray bars indicate the conditions GPR-GPC and OPR-OPC.

R-GPC (dark-gray bar), and 4.9 dB when defined as GPR-GPC (light-gray bar). Both CMR effects were highly significant [$t(5)=6.73$, $p=0.001$ and $t(5)=6.85$, $p=0.001$, respectively] and were not significantly different from one another [$t(5)=0.81$, $p=0.456$]. The results from the second control experiment, where the postcursors were presented at intermediate frequencies (off-frequency postcursors-OP) show slightly elevated thresholds both for the random (OPR) and the comodulated (OPC) conditions compared to the standard thresholds, R and C. However, the amount of CMR, as measured by the difference between OPR and OPC conditions [right-hand light-gray bar; 4.6 dB, $t(5)=4.26$, $p<0.01$], was highly significant and was close (within 1.5 dB) to that found in the standard condition without postcursors, R-C.

In summary, the results from the two additional control conditions indicate that the effects of the postcursors cannot be ascribed to either long-term, across-trial adaptation or to a general distraction or attentional effect produced by the presence of the postcursors. The findings support the idea that retroactive effects of perceptual segregation can lead to a deterioration in target detection.

IV. DISCUSSION

The results show that across-frequency modulation processing may interact with the processes that give rise to auditory object and stream formation. The effects of poststimulus manipulations make an interpretation based solely on neural inhibition or forward suppression (Las *et al.*, 2005) unlikely. Instead, the current results suggest the influence of higher-level processes, whereby modulations can be processed efficiently across different frequency regions only if they form part of the same auditory object. Stated another way, the modulation analysis observed in the tasks of the present study seems to be performed on objects, rather than frequency channels.

Retroactive effects in hearing, although rare, have been reported before. For instance, in speech perception, segments of a sound occurring after the offset of a vowel can affect the perceived identity of the vowel (Darwin, 1984; Darwin *et al.*, 1989; Roberts and Moore, 1990); noise bursts can be perceived differently depending on the following vowel (Liberman *et al.*, 1952); and certain features of sounds are perceived or remembered less well when followed by a masking sound (Massaro, 1975). Warren (1970) found strong retroactive effects of context on the recognition of "missing" speech sounds, when parts of a speech sound in recorded sentences were replaced with an extraneous sound (such as a tone or a gap). Retroactive effects on the simple detection of an auditory target are less commonly observed. One example is backward masking, which occurs when a brief target, e.g., a tone pulse, is presented just before a masker (e.g., Elliott, 1962; Oxenham and Moore, 1995). Another example is related to the detection of a brief target that is gated on synchronously with a masker, similar to the well-known "overshoot" effect (e.g., Zwicker, 1965). When the masker is gated on and off with the target, thresholds are often higher than when the masker continues beyond the offset of the target (Kidd and Wright, 1994). In this case, the additional masker energy *improves* performance, possibly by eliminating the potential masking produced by the masker offset transient. Such effects are typically observed only for very short target durations and for maskers that follow immediately (within 10 ms) of the end of the masker. The targets used in the present study were much longer (187.5 ms), as were the gaps between the flankers and the postcursors (62.5 ms), suggesting that the effects observed here are probably not related to those of backward masking or overshoot.

The results place strong constraints on the search for neural correlates of across-frequency modulation processing. The current findings seem incompatible with recent physiological studies suggesting that neural correlates of CMR may be found at the brainstem level, i.e., at an early stage of auditory processing (Pressnitzer *et al.*, 2001; Verhey *et al.*, 2003; Neuert *et al.*, 2004). In Pressnitzer *et al.* (2001), some of the recorded units in the cochlear nucleus (CN) of guinea pigs showed responses consistent with perceptual CMR. The addition of a comodulated flanking band in a CMR paradigm produced a strong reduction in the response to the masker band modulation, making the signal more salient in the corresponding poststimulus time histograms. A decision statistic based on d' showed that threshold was reached at lower signal levels for the comodulated condition than for the reference condition. Using a computational model, Pressnitzer *et al.* (2001) and Meddis *et al.* (2002) demonstrated that a

simple neural circuit consisting of the inhibition of a narrowband unit by a wideband inhibitor is able to replicate many of the physiological findings. These results thus provided evidence for an enhanced representation of the signal in the brainstem when presented in comodulated backgrounds. It is also possible that certain neurons in the brainstem that respond after a signal has ended show an altered representation when a subsequent masker is presented. However, the data shown in the present study lead us to question to what extent such across-channel processing actually reflects a direct physiological correlate of the perceptual across-channel CMR.

Several studies have investigated neural correlates of CMR at higher stages in the auditory pathways. In the primary auditory cortex of the cat, Nelken *et al.* (1999) considered the disruption of a neuron's envelope-following response as a correlate for CMR. Using a flanking-band experiment conceptually similar to the one considered in the present study, Nelken *et al.* (1999) showed that single units in the auditory cortex can demonstrate a response consistent with CMR. These units tended to lock to the envelope of the slowly fluctuating noise, whereas the addition of a low-level tone suppressed the envelope locking, a phenomenon that was referred to as "locking suppression." Such disruption of the envelope-following response to the masker at a cortical level was considered as a complementary strategy of achieving an enhanced signal representation in a modulated noise background (Nelken *et al.*, 1999; Langemann and Klump, 2001; Verhey *et al.*, 2003).

In a later study, the evolution of locking suppression along the auditory pathway was investigated (Las *et al.*, 2005). Recordings were made in the primary auditory cortex (A1) and in two preceding subcortical nuclei, the inferior colliculus (IC) and the medial geniculate body (MGB). Las *et al.* (2005) showed that, whereas responses in IC resembled in many aspects those in CN, a new response pattern appeared in MGB and became dominant in A1, whereby the representation of the tone was more explicit in the higher stages than in the brainstem. Specifically, Las *et al.* (2005) proposed that the enhancement of the representation of the low-level tone in slowly fluctuating noise-by suppression of envelope locking in the ascending auditory system-could be a correlate of the formation of an auditory object (the tone) as a separate entity from the background noise.

Nelken (2004) and Las *et al.* (2005) proposed that, while most of the physical attributes of the sound and many interesting auditory features might already be extracted in the brainstem (e.g., the IC), the *organization* of these features into auditory objects takes place in the auditory cortex using temporal and spectral contexts at several time scales. The nonlinear interaction between stimulus components in the primary auditory cortex thus results in a more abstract representation of sounds in terms of auditory objects. The perceptual data from the present study seem to be consistent with this interpretation. However, it remains unclear to what extent such a process may account for the grouping effects observed here. The decision mechanisms that can relate neural activity and percepts in scene analysis experiments need to be specified. Such mechanisms will need to operate over more than the duration of the stimulus, to account for the retroactive effects shown here. The stimulus configurations tested in the present study might provide a basis for explicitly testing the existing hypotheses regarding the neural representation of across-channel CMR.

Finally, the results of this study also provide challenges for future models of modulation processing and perception. While the recent model of Piechowiak *et al.* (2007), which reflects an across-channel extension of the modulation filterbank model of Dau *et al.* (1997), can account for a variety of detection and masking data, including across-channel CMR, it is not able to simulate the elimination of CMR as a result of perceptual segregation of the masker band from the flanker bands. Likewise, recent models of spectro-temporal processing in the auditory system proposed by Chi *et al.* (2005) cannot account for this finding. In both modeling approaches, the signal energy across time and frequency is essentially integrated linearly. This might be successful in relatively simple sound conditions but fails in more complex sound situations where perception depends on the acoustical context. However, the output of the processing models might provide some of the important auditory features as input to the "central processor." The models have been shown to be valuable as pre-processors in, for example, automatic speech recognition and objective assessment of speech quality (Hansen and Kollmeier, 1999; Tchorz and Kollmeier, 1999; Chi *et al.*, 2005). It remains to be seen how best such representations can be manipulated to predict the effects described in this study.

ACKNOWLEDGMENTS

This work was supported by the Danish Research Council (Ministeriet for Videnskab, Teknologi og Udvikling), the German Research Foundation (Deutsche Forschungsgemeinschaft), and the National Institutes of Health (Grant No. R01 DC 03909). The authors would like to thank Brian C. J. Moore, Daniel Pressnitzer, and one anonymous reviewer for their very helpful and supportive comments.

Altmann, C. F., Nakata, H., Noguchi, Y., Inui, K., Hoshiyama, M., Kaneoke, Y., and Kakigi, R. (**2007**). "Temporal dynamics of adaptation to natural sounds in the human auditory cortex," Cereb. Cortex **18**, 1350–1360.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organisation of Sound* (MIT, Cambridge, MA).

Bregman, A. S., and Pinker, S. (**1978**). "Auditory streaming and the building of timbre," Can. J. Psychol. **32**, 19–31.

Buss, E., and Hall, J. W. (**2008**). "Factors contributing to comodulation masking release with dichotic maskers," J. Acoust. Soc. Am. **124**, 1905–1908.

Buus, S. (**1985**). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958–1965.

Calford, M. B., and Semple, M. N. (**1995**). "Monaural inhibition in cat auditory cortex," J. Neurophysiol. **73**, 1876–1891.

Carlyon, R. P., Buus, S., and Florentine, M. (**1989**). "Comodulation masking release for three types of modulator as a function of modulation rate," Hear. Res. **42**, 37–46.

Chi, T., Ru, P., and Shamma, S. A. (**2005**). "Multiresolution spectrotemporal analysis of complex sounds," J. Acoust. Soc. Am. **118**, 887–906.

Cohen, M. F., and Schubert, E. D. (**1987**). "The effect of cross-spectrum correlation on the detectability of a noise band," J. Acoust. Soc. Am. **81**, 721–723.

Dannenbring, G. L., and Bregman, A. S. (**1978**). "Streaming vs. fusion of sinusoidal components of complex waves," Percept. Psychophys. **24**, 369–376.

Darwin, C. J. (**1984**). "Perceiving vowels in the presence of another sound: Constraints on formant perception," J. Acoust. Soc. Am. **76**, 1636–1647.
Darwin, C. J., Pattison, H., and Gardner, R. B. (**1989**). "Vowel quality changes produced by surrounding tone sequences," Percept. Psychophys. **45**, 333–342.
Dau, T., Ewert, S. D., and Oxenham, A. J. (**2005**). "Effects of concurrent and sequential streaming in comodulation masking release,"in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer, New York), pp. 335–343.
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**, 2893–2905.
Elliott, L. L. (**1962**). "Backward masking: Monotic and dichotic conditions," J. Acoust. Soc. Am. **34**, 1108–1115.
Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.
Grose, J. H., and Hall, J. W. (**1993**). "Comodulation masking release: Is comodulation sufficient?," J. Acoust. Soc. Am. **93**, 2896–2902.
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.
Hansen, M., and Kollmeier, B. (**1999**). "Continuous assessment of time varying speech quality," J. Acoust. Soc. Am. **106**, 2888–2899.
Kidd, G., and Wright, B. A. (**1994**). "Improving the detectability of a brief tone in noise using forward and backward masker fringes: Monotic and dichotic presentations," J. Acoust. Soc. Am. **95**, 962–967.
Klump, G. M. (**1996**). "Bird communication in the noisy world,"in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Comstock, Ithaca, NY), pp. 321–338.
Langemann, U., and Klump, G. M. (**2001**). "Signal detection in amplitude modulated maskers. I. Behavioural auditory thresholds in a songbird," Eur. J. Neurosci. **13**, 1025–1032.
Las, L., Stern, E. A., and Nelken, I. (**2005**). "Representation of tone in fluctuating maskers in the ascending auditory system," J. Neurosci. **25**, 1503–1513.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Liberman, A. M., Delattre, P., and Cooper, F. S. (**1952**). "The role of selected stimulus-variables in the perception of the unvoiced stop consonants," Am. J. Psychol. **65**, 497–516.
Massaro, D. W. (**1975**). "Backward recognition masking," J. Acoust. Soc. Am. **58**, 1059–1065.
McFadden, D. (**1986**) "Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band," J. Acoust. Soc. Am. **80**, 1658–1667.
McFadden, D., and Wright, B. A. (**1992**). "Temporal decline of masking and comodulation masking release," J. Acoust. Soc. Am. **92**, 144–156.
Meddis, R., Delahaye, R., O'Mard, L., Summer, C., Fantini, D. A., Winter, I., and Pressnitzer, D. (**2002**). "A model of signal processing in the cochlear nucleus: Comodulation masking release," Acta. Acust. Acust. **88**, 387–398.
Nelken, I. (**2004**). "Processing of complex stimuli and natural scenes in the auditory cortex," Curr. Opin. Neurobiol. **14**, 474–480.
Nelken, I., Rotman, Y., and Yosef, O. B. (**1999**). "Responses of auditory cortex neurons to structural features of natural sounds," Nature (London) **397**, 154–157.
Neuert, V., Verhey, J. L., and Winter, I. M. (**2004**). "Responses of dorsal cochlear nucleus neurons to signals in the presence of modulated maskers," J. Neurosci. **24**, 5789–5797.
Oxenham, A. J., and Moore, B. C. J. (**1995**). "Additivity of masking in normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **98**, 1921–1934.
Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.
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, New York), pp. 123–178.
Piechowiak, T., Ewert, S. D., and Dau, T. (**2007**). "Modeling comodulation masking release using an equalization-cancellation mechanism," J. Acoust. Soc. Am. **121**, 2111–2126.
Pressnitzer, D., Meddis, R., Delahaye, R., and Winter, I. M. (**2001**). "Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus," J. Neurosci. **21**, 6377–6386.
Roberts, B., and Moore, B. C. J. (**1990**). "The influence of extraneous sounds on the perceptual estimation of the first format frequency in vowels," J. Acoust. Soc. Am. **88**, 2571–2583.
Schooneveldt, G. P., and Moore, B. C. J. (**1987**). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. **82**, 1944–1956.
Tchorz, J., and Kollmeier, B. (**1999**). "A model of auditory perception as front end for automatic speech recognition," J. Acoust. Soc. Am. **106**, 2040–2050.
Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (**2004**). "Multiple time scales of adaptation in auditory cortex neurons," J. Neurosci. **24**, 10440–10453.
Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," J. Acoust. Soc. Am. **106**, 2733–2745.
Verhey, J. L., Pressnitzer, D., and Winter, I. M. (**2003**). "The psychophysics and physiology of comodulation masking release," Exp. Brain Res. **153**, 405–417.
Warren, R. M. (**1970**). "Restoration of missing speech sounds," Science **167**, 392–393.
Wehr, M., and Zador, A. M. (**2005**). "Synaptic mechanisms of forward suppression in rat auditory cortex," Neuron **47**, 437–445.
Zwicker, E. (**1965**). "Temporal effects in simultaneous masking by white-noise bursts," J. Acoust. Soc. Am. **37**, 653–663.

Can temporal fine structure represent the fundamental frequency of unresolved harmonics?

Andrew J. Oxenham,[a] Christophe Micheyl, and Michael V. Keebler
Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

(Received 21 March 2008; revised 4 February 2009; accepted 4 February 2009)

At least two modes of pitch perception exist: in one, the fundamental frequency (F0) of harmonic complex tones is estimated using the temporal fine structure (TFS) of individual low-order resolved harmonics; in the other, F0 is derived from the temporal envelope of high-order unresolved harmonics that interact in the auditory periphery. Pitch is typically more accurate in the former than in the latter mode. Another possibility is that pitch can sometimes be coded via the TFS from *unresolved* harmonics. A recent study supporting this third possibility [Moore *et al.* (2006a). J. Acoust. Soc. Am. **119**, 480–490] based its conclusion on a condition where phase interaction effects (implying unresolved harmonics) accompanied accurate F0 discrimination (implying TFS processing). The present study tests whether these results were influenced by audible distortion products. Experiment 1 replicated the original results, obtained using a low-level background noise. However, experiments 2–4 found no evidence for the use of TFS cues with unresolved harmonics when the background noise level was raised, or the stimulus level was lowered, to render distortion inaudible. Experiment 5 measured the presence and phase dependence of audible distortion products. The results provide no evidence that TFS cues are used to code the F0 of unresolved harmonics. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3089220]

PACS number(s): 43.66.Hg, 43.66.Nm, 43.66.Ba [RLF] Pages: 2189–2199

I. INTRODUCTION

Many sounds in our environment, such as voiced speech, musical tones, and some animal vocalizations, are harmonic, comprising frequencies that are all at, or close to, integer multiples of a fundamental frequency (F0). We tend to hear a pitch corresponding to the F0, even when there is no energy at the F0 itself. This phenomenon has various terms, including the "pitch of the missing fundamental," periodicity pitch, and residue pitch (e.g., Schouten, 1940; Licklider, 1954).

The mechanisms of pitch perception have been the subject of numerous studies over the past century and are still being debated today (Plack *et al.*, 2005). There is broad consensus on certain aspects of pitch perception. For instance, low-numbered harmonics (<10) typically produce a more salient pitch and more accurate F0 discrimination than do high-numbered harmonics. The transition from strong to weak pitch with increasing lowest harmonic number (N) within a complex corresponds reasonably well with the transition from resolved to unresolved harmonics, as estimated by the dependence of F0 difference limens (F0DLs) on the phase relationships between components (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2006a, 2006b), although it does not appear to be peripheral resolvability per se that determines the changes in percept with increasing N (Houtsma and Goldstein, 1972; Bernstein and Oxenham, 2003, 2008).

Resolved harmonics (such as isolated pure tones) may be coded by their tonotopic (place) representation (e.g., Wightman, 1973), by phase-locking to the temporal fine structure (TFS) in the auditory nerve (e.g., Meddis and O'Mard, 1997), or by a combination of both (Shamma and Klein, 2000; Oxenham *et al.*, 2004). In so-called "pattern recognition" models of pitch, the estimates of the individual frequencies are combined to derive the overall F0 (Goldstein, 1973; Terhardt, 1974).

For unresolved harmonics, which interact within the passband of single peripheral auditory filters, the F0 may be extracted via phase-locking to the temporal envelope of the complex waveform after peripheral filtering, by phase-locking to TFS peaks located near envelope peaks, or both. Early evidence in favor of the use of TFS to derive the F0 from unresolved harmonics came from experiments using sinusoidally amplitude-modulated (SAM) tones (e.g., de Boer, 1956; Schouten *et al.*, 1962). A SAM tone has three tonal components, the carrier frequency (f_c) and two side bands, one above and one below the carrier, with the frequency spacing between the components corresponding to the modulation frequency (f_m). When f_c is an integer multiple of f_m, the waveform is periodic and consists of three consecutive harmonics with an F0 equal to f_m. When f_c is not an integer multiple of f_m, the temporal envelope is still periodic, with a frequency f_m; however the TFS no longer shares the same F0. The fact that listeners heard a shift in pitch when the f_c was shifted, even when f_m remained the same, was taken as evidence that listeners were sensitive to the TFS of complex waveforms (de Boer, 1956; Schouten *et al.*, 1962). One potential confound in these early experiments was that the spectral centroid of the complex shifted with the f_c, leaving open the possibility that listeners were responding to a change in the spectrum of the three-tone complex rather than to a change in the TFS. Moore and Moore (2003) ad-

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

dressed this issue by using stimuli with more components, which were either harmonically related or were all shifted upward by the same amount in Hertz, and filtering them such that the spectral envelope of the stimuli remained constant. When the lowest harmonic number (N) present was about 14 or higher, no pitch shifts were heard when the frequencies of the components were shifted, suggesting that listeners were only sensitive to the temporal envelope and not to the TFS. At low values of N a pitch shift was heard, but this may have been due to the frequency shifts of the individual resolved harmonics. At intermediate N values of around 9, significant pitch shifts were found. Moore and Moore (2003) concluded that the results at intermediate values of N could be interpreted in two ways: First, if the harmonics were partially resolved, listeners may have been able to extract the individual frequencies of some of the harmonics and thus perceived a pitch shift by way of a shift in the individual harmonic frequencies; second, if the harmonics were unresolved, listeners may have been sensitive to the TFS of the complex waveform, in line with the theories of de Boer (1956) and Schouten *et al.* (1962). In a follow-up study, Moore *et al.* (2006a) attempted to distinguish between these two possibilities.

Moore *et al.* (2006a) measured F0DLs for three-component harmonic complexes centered at 2000 Hz as a function of F0 or, equivalently, the lowest harmonic number present. Either the three components were all in cosine phase (COS) or the middle component was shifted by 90° to produce what is often referred to as alternating phase (ALT). It was assumed that accurate F0 discrimination, as reflected by low F0DLs, implied that listeners were able *either* to access resolved harmonics *or* to process the TFS information from unresolved harmonics. Moore *et al.* (2006a) proposed to distinguish between these two possibilities by assessing whether or not F0DLs were dependent on the component phase relationships: if the components were resolved, then F0DLs should be independent of phase, whereas if they were unresolved, the components would interact within the auditory periphery and F0DLs might be phase-dependent. When N was 6 or 7, Moore *et al.* (2006a) found that F0DLs were low, and there was no effect of component phase, consistent with the harmonics being resolved. When N was 8, mean F0DLs in the COS condition were still as low as those when N was 6 or 7, but now mean F0DLs in the ALT condition were about a factor of 2 higher. The combination of low F0DLs and a phase effect led Moore *et al.* (2006a) to conclude that listeners were indeed using TFS to code the F0 from unresolved harmonics for values of N between about 8 and 10.

The conclusions of Moore *et al.* (2006a) have important theoretical implications. They suggest that listeners are able to use TFS for extracting not only information about the frequencies of individual resolved harmonics or the center frequency of a narrowband sound but also information about the F0 of a complex tone. Although the idea stems from early pitch research (e.g., Schouten, 1940; de Boer, 1956), more recent approaches have tended to consider pitch perception as originating from two cues [or possibly two mechanisms—see Carlyon and Shackleton (1994)], one that involves the individual frequencies of resolved harmonics (coded via place and/or TFS) and one that involves the F0 of the unresolved harmonics (coded via temporal envelope cues) (e.g., Shamma and Klein, 2000). The suggestion that TFS may play another role for unresolved harmonics in the "intermediate" range of harmonic numbers (for N between 8 and 11) can be interpreted as adding a third pitch mechanism, which has implications for models of pitch processing. Another important implication of the results of Moore *et al.* (2006a) relates to the current debate regarding the role of TFS encoding in relation to the listening difficulties experienced by hearing-impaired individuals. Recent studies have extended the work of Moore *et al.* (2006a) to hearing-impaired listeners (Moore *et al.*, 2006b; Hopkins and Moore, 2007) and have shown that in most of these listeners, there was no evidence for this intermediate region, where phase effects are found despite good (low) F0DLs. This has been interpreted as evidence that hearing-impaired listeners have reduced access to TFS cues (Moore *et al.*, 2006b; Hopkins and Moore, 2007). The idea that hearing-impaired listeners have difficulties extracting TFS information in a way that cannot be simply explained by poorer frequency selectivity (and fewer resolved harmonics) has also been extended to studies of speech reception (e.g., Lorenzi *et al.*, 2006; Hopkins *et al.*, 2008). Because of their theoretical importance and their influence on subsequent research on TFS coding, the results of Moore *et al.* (2006a) deserve closer consideration.

One puzzling aspect of the results of Moore *et al.* (2006a) is that they found a strong phase effect even when there was no increase in the F0DLs for the COS complex. Although this is crucial to their interpretation that TFS coding is involved, it does not seem consistent with earlier studies. For instance, Houtsma and Smurzynski (1990) used complexes that were either in sine phase, which is thought to produce a highly modulated temporal envelope after auditory filtering, or in negative Schroeder phase (Schroeder, 1970), which is thought to produce a much less modulated temporal envelope (e.g., Kohlrausch and Sander, 1995; Oxenham and Dau, 2001). Houtsma and Smurzynski (1990) found that phase affected performance only once F0DLs in the sine-phase complex were elevated. Similar results were obtained in studies by Bernstein and Oxenham (2005, 2006b) when comparing sine-phase and random-phase complexes. Moore *et al.* (2006a) suggested that the difference might be due to the possibility that random and Schroeder phases do not produce the "optimally" flat temporal envelope produced by ALT-phase complexes and that the earlier studies had used complexes with more components.

Another difference between the study of Moore *et al.* (2006a) and the previous ones is the level of background noise used to mask distortion products. Houtsma and Smurzynski (1990) used a pink noise and presented their tones 20 dB above masked threshold; Bernstein and Oxenham (2003, 2005, 2006b) used a noise that produced roughly equal pure-tone detection thresholds at all frequencies (e.g., Moore *et al.*, 2000), and presented their tones on average between 10 and 15 dB above masked threshold. In contrast, Moore *et al.* (2006a) used threshold equalizing noise at a level of 30 dB/ERB_N, where ERB_N refers to the average value of the equivalent rectangular bandwidth of the auditory

filter for young normal-hearing listeners at moderate levels (Glasberg and Moore, 1990). Because their complexes were presented at a level of 60.2 dB SPL (sound pressure level) per component, the tones were likely to have been 30–35 dB above their masked threshold.

The level of distortion products induced by two-, three-, and multitone complexes has been the subject of much research (e.g., Goldstein, 1967; Smoorenburg, 1972a, 1972b; Buunen *et al.*, 1974; Pressnitzer and Patterson, 2001). The effective level of the distortion products is influenced by a number of variables, but it is not unusual to find distortion products at levels as high as 20 dB below the level of the primary components. The level and phase of distortion products have been shown to vary with the phase relations of the (three or more) primary components (Buunen *et al.*, 1974; Pressnitzer and Patterson, 2001), and a number of researchers have proposed that some phase effects found in pitch perception can be ascribed to changes in the levels of distortion products (Goldstein, 1973; Buunen *et al.*, 1974; Fleischer, 1976; for an early review, see Moore, 1977). Thus, it appears that the noise level used by Moore *et al.* (2006a) may not have been sufficient to mask distortion products. This in turn implies that changes in the relative levels of the distortion products with changes in phase may have influenced their results. The present study was designed to repeat the study of Moore *et al.* (2006a) with the same level of noise used in the original study and with a higher level of background noise to assess the extent to which their results and conclusions were affected by the audibility of distortion products.

II. EXPERIMENT 1: REPLICATION WITH LOW BACKGROUND NOISE LEVEL

A. Methods

1. Subjects

Four listeners (aged 18–22 yr) took part. All four had normal hearing, defined as having audiometric thresholds of 20 dB HL (hearing level) or less at octave frequencies between 250 and 8000 Hz. Following pure-tone audiometry, the listeners were given the opportunity to familiarize themselves with the stimuli and task. In all experiments in this study, all listeners had some musical education and had played a musical instrument at some point in their life. In addition, most of them had already participated in pitch discrimination experiments prior to this study. Therefore, they had no difficulty understanding the instructions, and by the end of the first 2-h session, their thresholds already fell within the same range as those obtained by one of the authors, who had extensive experience in pitch discrimination tasks, and they showed no clear signs of further improvement. In this and all subsequent experiments, elevated thresholds, obtained at first in a few of the listeners (mostly those with the least amount of musical training or prior experience in pitch discrimination tasks), were discarded before actual data collection began.

2. Procedure

This experiment measured F0DLs using a two-interval two-alternative forced-choice method with a 3-down 1-up adaptive procedure that tracks the 79% correct point on the psychometric function (Levitt, 1971). The two intervals contained complex tones with F0s that differed by an amount of ΔF0, expressed as a percentage of the F0 around which the two interval F0s were geometrically centered ($F0_c$). The initial value of ΔF0 was 20%, which was initially varied (increased or decreased, according to the adaptive procedure) by a factor of 1.414 for the first four reversals in the adaptive rule. The factor was reduced to 1.189 for the final four reversals. Threshold was defined as the geometric mean value of ΔF0 at the last four reversals. At least four threshold estimates were obtained for each subject in each condition. The reported threshold was taken as the geometric mean of all estimates.

3. Stimuli

The stimuli were the same as those used by Moore *et al.* (2006a). Each complex tone consisted of three consecutive harmonics, each with a level of 60.2 dB SPL (65 dB SPL overall). The nominal number of the lowest harmonic, N, ranged from 4 to 14. The nominal frequency of the center component was 2000 Hz, which was roved by $\pm 10\%$ on each trial to encourage listeners to base their judgments on comparisons within each trial, rather than on any long-term memory representations. To reduce the effectiveness of spectral cues (as opposed to F0 cues), the value of N was roved across intervals, such that the actual lowest harmonic for each stimulus could be $N-1$, N, or $N+1$. The roved value was selected independently (with replacement) in each interval. The three components were added either in cosine phase (COS condition) or in alternating phase (ALT condition), where the phase of the center component lagged by 90° (sine phase). Each complex tone had a total duration of 480 ms, gated on and off with 20-ms raised-cosine ramps. The two complex tones within each trial were separated by an interstimulus interval of 300 ms.

A background threshold equalizing noise (Moore *et al.*, 2000) was added, which was gated on 400 ms before the first interval and gated off 400 ms after the second interval. The noise was generated in the spectral domain and contained energy between 50 and 3000 Hz. The level of the noise was set to 30 dB SPL/ERB_N, so that the individual tones were about 30–35 dB above their masked thresholds in the noise.

B. Results and discussion

The mean results across the four listeners are shown in Fig. 1, where F0DLs are plotted as a function of the lowest harmonic number (N) of the three-tone complex for COS (open symbols) and ALT (filled symbols) complexes. Despite inter-individual variability, as also found by Moore *et al.* (2006a), the mean data obtained in this experiment replicated the main findings from the experiment of Moore *et al.* (2006a): for low values of N, F0DLs were low and similar in COS and ALT phases; at high values of N, F0DLs were higher and were generally higher in ALT than in COS phase.

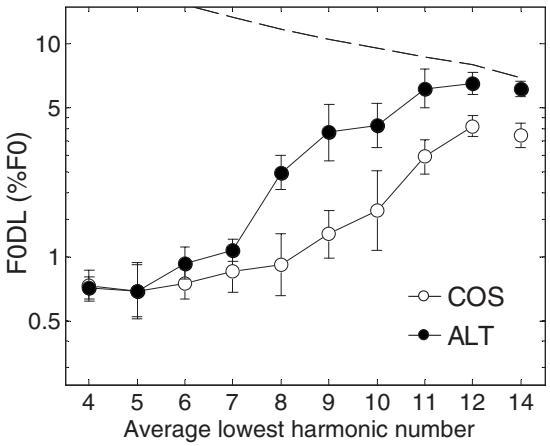


FIG. 1. Mean F0DLs for three-component complexes as a function of the lowest harmonic number present (N). Filled symbols represent results for stimuli in ALT phase; open symbols represent results for stimuli in COS phase. Error bars represent ±1 standard error of the mean. The level of the threshold equalizing noise was set at 30 dB/ERB$_N$, meaning that the components were about 30–35 dB above their masked thresholds in noise. The dashed line shows the predicted thresholds based solely on spectral, rather than F0, information.

According to the reasoning of Moore *et al.* (2006a), if TFS can be used to code the F0 of unresolved harmonics, then some conditions should exist where (a) F0DLs are small, implying accurate coding using TFS, and (b) a phase effect is found, implying that the harmonics are unresolved. Such conditions can be seen in Fig. 1 for values of N between 8 and 10, where F0DLs are still relatively small in the COS condition but not in the ALT condition. To test this more formally, we established a "cutoff" F0DL, at the midpoint between the "asymptotic" low and high F0DLs, estimated using the geometric mean of the F0DLs in the COS condition at the lowest and highest value of N tested. We then took the highest value of N at which the F0DL in the COS condition fell below the cutoff value for the F0DL; this represented the condition for which the harmonics were most likely to be unresolved and still produce low F0DLs. Finally, we performed a paired t-test comparing the (log-transformed) F0DLs in the COS and ALT conditions for the selected value of N.

For the data in Fig. 1, the cutoff F0DL value was 1.64%. The highest N value for which the mean COS F0DLs fell below the cutoff was 10. A paired t-test at this value of N produced a significant effect of phase [paired $t(3)=2.93$, one-tailed $p=0.031$]. This finding of significant phase interactions for conditions in which the F0DLs are low is consistent with the results and conclusions of Moore *et al.* (2006a). In other words, these data show that a phase effect can occur, even when F0DLs are low, in line with the idea that TFS can be used to code the F0 of unresolved harmonics.

As the value of ΔF0 increases, the roving of lowest harmonic number by ±1 ceases to rule out the use of spectral (as opposed to F0) cues. The dashed curve in Fig. 1 shows the prediction of performance based solely on the (perfect or noiseless) frequency discrimination of the lowest component in each complex, as described in detail by Moore *et al.* (2006a). It can be seen that the mean F0DLs at the highest values of N approached this limit in the ALT, but not the COS, condition.

C. Experiment 1a: Using a different roving technique

As discussed by Moore *et al.* (2006a), their roving technique (also used here in experiment 1) leads to a listener obtaining 67% correct responses at small values of ΔF0, even if all responses are based solely on the increase in the frequency of the lower spectral edge of the stimulus and not on the change in F0. This is because with the three possible values of the lowest harmonic number in each interval (N_{INT}), the probability that the N of the interval with the higher F0 will be the same as or greater than the N of the interval with the lower F0 is 67% (or 6 out of 9). Although 67% is lower than the tracking percent correct of 79%, it is still considerably higher than the 50% value of chance in a more typical two-alternative forced-choice experiment. Also, for two out of the nine possible combinations of N, the difference in N between the intervals was 2, producing a very large change in timbre, which listeners may have found distracting. Both these factors may have led subjects to lend more weight to the spectral pitch of the stimuli, rather than the desired periodicity pitch. Experiment 1a was designed to avoid these potential confounds by restricting the difference in N_{INT} between the two intervals within a trial to 1 and by ensuring that at small values of ΔF0, listeners would only achieve 50% correct by using spectral edge cues. We achieved this goal by ensuring that the lowest harmonic number in the two intervals was always different and always differed by only 1. In other words, the possible combinations of the lowest harmonic number in the two intervals were $(N-1,N)$, $(N,N-1)$, $(N,N+1)$, and $(N+1,N)$. All other details in the experiment were the same as in the main experiment, except that the background noise was now broadband (extending from 50 Hz to 19.2 kHz) instead of being lowpass filtered at 3 kHz. Seven normal-hearing subjects took part in this experiment. Two of these listeners also took part in experiment 1. The ages of the subjects ranged from 19 to 33. The different roving paradigm led to some different predictions for thresholds based only on the frequency of the lowest harmonic. Briefly, the frequency of the lowest harmonic is a consistently reliable cue when $(N-1)\mathrm{F0}_H > N\mathrm{F0}_L$, where $\mathrm{F0}_H$ and $\mathrm{F0}_L$ are the F0s of the higher- and lower-F0 interval within a trial, respectively. Given that $\mathrm{F0}_H=\mathrm{F0}_L(1+\Delta\mathrm{F0}/100)$, where ΔF0 is the difference between $\mathrm{F0}_H$ and $\mathrm{F0}_L$ in percent, the inequality can be solved for ΔF0, such that spectral cues are reliable under the following condition:

$$\Delta\mathrm{F0} > 100[N/(N-1)-1]. \quad (1)$$

This prediction is shown as a dashed curve in Fig. 2, along with the mean results from experiment 1a. It can be seen that F0DLs were somewhat higher overall than in experiment 1. This is consistent with the prediction that a higher sensitivity (d') is required to achieve 79% correct when performance based on spectral cues is 50%, as opposed to 67%, although part of the difference may simply be due to inter-subject

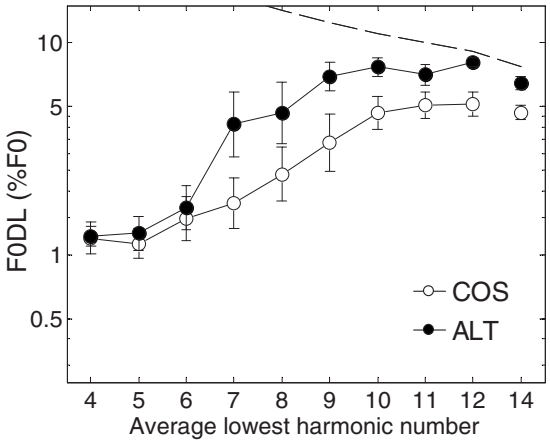


FIG. 2. As in Fig. 1, except that the noise was broadband, instead of low-pass filtered at 3 kHz, and that the roving of the N followed a different procedure (see text for details). As in Fig. 1, error bars in this and following figures represent ± 1 standard error of the mean. In this and all following figures, the dashed line shows predicted thresholds based on spectral rather than F0 information. These predictions differ slightly from those shown in Fig. 1 due to the use of a different lower-harmonic-number randomization rule.

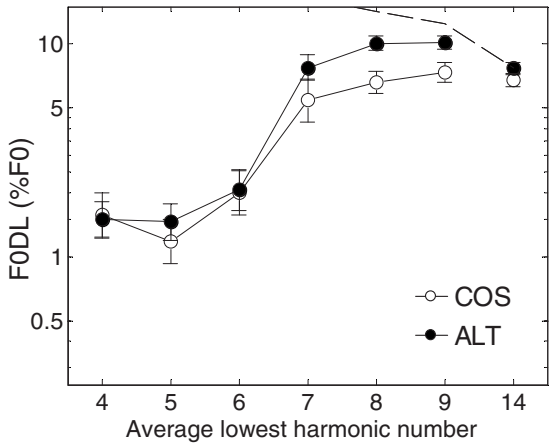


FIG. 3. As in Fig. 2, except that the level of the background noise was 20 dB higher, which should have been sufficient to mask distortion products.

variability. Other than that, the pattern of results was similar to that found in the main experiment.

Using the same analysis that was used for experiment 1, the cutoff F0DL (i.e., the geometric mean of the lowest and highest group F0DLs) was 2.37%, and the highest value of N with an F0DL below that cutoff in the COS condition was 7. For this value of N, a paired t-test revealed a significant effect of phase [paired $t(6)=2.92$, one-tailed $p=0.014$]. Thus, consistent with the results from experiment 1, conditions existed in which F0DLs were low but a phase effect was observed. The similarity in the pattern of results between experiments 1 and 1a was supported by a mixed-model analysis of variance (ANOVA), with experiment as a between-subjects variable and N and phase as within-subject variables.[1] The ANOVA revealed a significant main effect of experiment [$F(1,9)=8.44$, $p=0.017$], in line with F0DLs being somewhat higher in experiment 1a, but no interactions of the other variables (N and phase) with experiment ($p>0.1$ in all cases).

III. EXPERIMENT 2: REDUCING THE AUDIBILITY OF DISTORTION PRODUCTS WITH BROADBAND NOISE

A. Rationale

Both experiments 1 and 1a were successful in replicating the basic results of Moore *et al.* (2006a), in particular the finding that low F0DLs, implying TFS processing, were found in conjunction with phase dependencies, implying unresolved harmonics, at intermediate values of N. However, as discussed in Sec. I, the background noise level was not sufficient to rule out the audibility of distortion products, the amplitudes of which may have varied with the phase relationships and may have influenced the results. This possibility was tested here by raising the level of the background noise to 50 dB SPL/ERB$_N$, which led to the individual components of the complex being between 10 and 15 dB above their masked thresholds.

B. Methods

The stimuli and procedure were the same as in experiment 1a, with the exception that the level of the background noise was increased by 20 dB. The roving paradigm was the same as that used in experiment 1a, so that at small values of ΔF0 listeners would obtain only 50% correct by basing their judgments on the frequency of the lowest harmonic present, rather than on the F0. Nine normal-hearing listeners (aged 18–33) took part in this experiment. Seven of them also participated in either experiment 1 or experiment 1a, or both. The experiments were not run systematically in the order in which they are described in this article. In fact, those listeners who took part in both experiment 2 and experiment 1 or 1a had completed experiment 2 first.

C. Results and discussion

The mean F0DLs across the nine listeners are plotted in Fig. 3. Again, filled symbols represent thresholds for ALT phase, and open symbols represent thresholds for COS phase. The dashed line represents predictions based solely on the frequency of the lowest harmonic, as described in experiment 1a. Many trends in the data are similar to those observed in experiments 1 and 1a. In particular, F0DLs are low at low values of N and increase at higher values of N, and differences between COS and ALT conditions only emerge at higher values of N. However, some differences between these and the earlier results are also apparent. Most importantly, for our purposes, is the apparent lack of a phase effect at values of N for which F0DLs were low. Using the same analysis that was used for experiment 1, the cutoff F0DL (i.e., the geometric mean of the lowest and highest group F0DLs) was 3.40%, and the highest value of N with an F0DL below that cutoff in the COS condition was 6. In contrast to experiments 1 and 1a, no significant phase effect was found for this value of N [paired $t(8)=0.219$, one-tailed $p=0.416$].

This pattern is not consistent with the results of Moore *et al.* (2006a), as replicated in experiments 1 and 1a, but it is

consistent with the earlier studies of Houtsma and Smurzynski (1990) and Bernstein and Oxenham (2005, 2006a). It seems that increasing the level of the noise to ensure that distortion products were masked was sufficient to eliminate the phase effects when F0DLs were small. The results therefore suggest that the phase effects found by Moore *et al.* (2006a) in conjunction with small F0DLs may have been produced by distortion products, rather than by the coding of F0 via the TFS of unresolved harmonics.

At least one puzzle remains. In the studies of Houtsma and Smurzynski (1990) and Bernstein and Oxenham (2003), the transition between good and poor performance occurred for N values between about 9 and 12, whereas in the present study the transition occurs between N values of 6 and 7. One possible reason is that the present study used only three harmonics, whereas the earlier studies used 12 or 13 consecutive harmonics. At face value, the larger number of harmonics does not provide a very satisfactory explanation if (as often assumed) it is the lowest harmonics present that primarily determine performance. On the other hand, the larger number of harmonics may provide more spectral (and timbral) stability, making the rove of the lowest harmonic less perceptually distracting. Another possibility relates to the range of F0s tested. In the present experiment, when $N=6$, the nominal F0 of the complex was about 286 Hz. This is higher than the F0s used by either of the earlier studies. In fact, a more recent study by Oxenham and Keebler (2007) found that even when 12 harmonics were present, the transition from good to poor performance occurred for values of N between 9 and 12 only when the F0 was either 100 or 200 Hz; for F0s of 300 Hz or higher, the transition occurred for values of N between 6 and 9, more in line with the current results. Thus, it may be that the apparent discrepancy between the present results and those of earlier studies is related more to the underlying F0 than to the number of harmonics presented. This conjecture should be explicitly tested by studying the effects of the number of harmonics present on the transition point between low and high F0DLs.

IV. EXPERIMENT 3: REDUCING THE AUDIBILITY OF DISTORTION PRODUCTS WITH LOWPASS NOISE

A. Rationale

The results of experiment 2 showed that the effect of phase on F0DLs only became apparent when F0DLs were poor, consistent with the idea that unresolved harmonics are coded by the temporal envelope and not TFS. It was suggested that the effect of the background noise was to mask distortion products. Another possibility is that the higher noise level used in experiment 2 interfered with the ability of listeners to use the TFS cues, via some form of direct masking of the stimulus components, rather than the distortion products. The aim of the present experiment was to test this hypothesis by using a lowpass-filtered noise with a cutoff frequency below the frequency of the lowest component present. The predictions were as follows: if the effect of the noise was to interfere with TFS processing, then reducing the on-frequency noise by lowpass filtering should result in an improvement in performance and a restoration of the phase effects observed by Moore *et al.* (2006a); on the other hand, if the main effect of the noise was to mask the distortion products, then lowpass filtering the noise should have little effect on the pattern of results, so long as the cutoff frequency is set such that the noise still masks the distortion products.

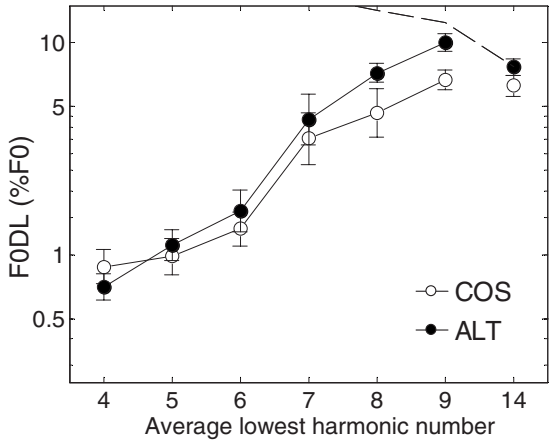


FIG. 4. As in Fig. 3, except that the noise was lowpass filtered with a cutoff frequency below the stimulus components but above the frequency of the likely distortion products.

B. Methods

The stimuli and procedure were the same as those used in experiment 2, with the exception that the high-frequency cutoff of the background noise was varied from trial to trial and was set such that it was 0.5F0 Hz lower than the lowest component presented in the two intervals of each trial. In this way, the noise energy did not directly overlap with any of the stimulus frequency components, but the cutoff frequency was always higher than the nearest distortion product in the spectrally lower of the two complexes (corresponding to the $2f_1$-f_2 distortion product, where f_1 and f_2 are the frequencies of the lower two components of the complex). Nine normal-hearing listeners participated in this experiment (aged between 18 and 22 yr); four of the subjects had also taken part in at least one of the previous experiments.

C. Results and discussion

The mean results across the nine listeners are shown in Fig. 4. Thresholds were somewhat lower overall than in experiment 2, suggesting that removing the "on-frequency" background noise had some effect. However, the overall pattern of results was very similar to that found in experiment 2 (Fig. 3), with low F0DLs for values of N up to 6 and higher F0DLs beyond that.

The search for low F0DLs in conjunction with significant phase effects was carried out in the same way as was done in experiments 1 and 2. The cutoff F0DL was found to be 2.42%, and the highest value of N for which the COS F0DL fell below the cutoff was again 6. At this value of N a paired t-test revealed no significant effect of phase [$t(8)=1.63$, one-tailed $p=0.071$].

Overall, neither experiment 2 nor experiment 3 provides support for the idea that phase effects can occur when F0DLs are small if potential distortion products are masked.

V. EXPERIMENT 4: F0DLS FOR COMPLEXES AT LOW SENSATION LEVELS

A. Rationale

This experiment provided a further test of whether the background noise was interfering with TFS processing, rather than just masking distortion products, by presenting the stimuli in quiet. To reduce the possibility that listeners could detect and use distortion products, the stimuli were presented at the relatively low sensation level of 20 dB above absolute threshold (20 dB SL). This low level should be sufficient to render distortion products inaudible but should nevertheless be sufficient for good pitch perception. For instance, Hoekstra (1979) showed that F0DLs for complex tones in noise approach an asymptotic value at around 20 dB above masked threshold.

B. Methods

The stimuli and procedure were the same as those used in experiment 2, with the following two differences: (1) there was no background noise present; (2) the three-tone complexes were presented at a level 20 dB above detection threshold in quiet. Detection thresholds in quiet were measured for three-component complexes, with the frequency of the middle component fixed at 2 kHz and the lowest harmonic number set to 6, 8, or 10 (corresponding to F0s of 285.71, 222.22, and 181.82 Hz). The overall level of the complex was varied adaptively, following a 3-down 1-up rule, which tracked thresholds corresponding to 79.4% correct. After thresholds had been estimated by taking the average of three adaptive runs for each subject for each of the three test conditions, the level of the test stimuli was set to 20 dB above the estimated threshold value for each subject individually. Five normal-hearing listeners participated in this experiment (aged between 18 and 22 yr). All of these subjects had also taken part in at least one of the previous experiments.

C. Results and discussion

The results are shown in Fig. 5. Despite the lack of background noise and the lower absolute level, the pattern of results was very similar to that found in experiments 2 (Fig. 3) and 3 (Fig. 4). In this case, the cutoff F0DL was 3.92%, and the highest value of N for which the COS F0DL fell below the cutoff was again 6. At this value of N a paired t-test revealed no significant effect of phase [paired $t(4)=0.49$, one-tailed $p=0.325$]. Overall, the results from this experiment in the absence of background noise were consistent with those from experiments 2 and 3, suggesting that the lack of a phase effect when F0DLs are low cannot be ascribed to direct masking or interference by the noise. A parsimonious interpretation of the results from all five experiments so far is that distortion products can affect F0DLs when audible (experiments 1 and 1a); when distortion products are not audible, no effects of phase are observed in conditions producing low F0DLs.

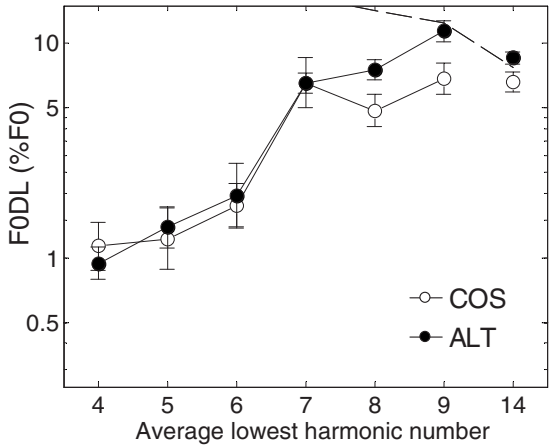


FIG. 5. As in Fig. 3, except that no background noise was present and that the complexes were presented at 20 dB SL.

All three experiments in which distortion products were deemed inaudible (experiments 2–4) yielded very similar patterns of results. The trends observed between and within each of three experiments were further studied using a mixed-model ANOVA, with experiment as a between-subjects variable and N and phase as within-subject variables.[2] The results of this analysis revealed a significant main effect of phase [$F(1,20)=24.73$, $p<0.0005$], a significant main effect of harmonic number [$F(5.38,107.76)=123.48$, $p<0.0005$], and a significant interaction between these two factors [$F(4.51,90.24)=5.02, p=0.001$]. *Post hoc* comparisons (paired t-tests) between COS and ALT thresholds pooled across all three experiments for each value of N showed a phase effect emerging at $N=8$ [$t(22)=5.25$, one-tailed $p<0.005$; for higher N's: $4.20<t(22)<14.66$, one-tailed $p<0.005$] but not below [for $4\leq N\leq 7$: $-2.00<t(22)<2.01$, one-tailed $p\geq 0.2$].[3] This pattern of results is consistent with the hypothesis that when distortion products are inaudible, phase effects are only observed in conditions where F0DLs are poor and are likely to be determined by temporal envelope cues. All three experiments showed a significant phase effect emerging at $N=8$, rather than at $N=7$, even though F0DLs were already high for $N=7$ (see Figs. 3–5). This outcome may be related to how resolvability affects pitch and phase effects differently: for tones to be unresolved, at least two must interact, but for a phase effect to be observed between tones that are not in an octave relationship, at least three components must interact. This might account for why a phase effect emerged at a value of N that was 1 higher than the value of N for which F0DLs became poor.

VI. EXPERIMENT 5: EFFECTS OF PHASE ON THE RELATIVE LEVEL OF DISTORTION PRODUCTS

A. Rationale

The results of the current study suggest that distortion products may play an important role in explaining the results

of Moore *et al.* (2006a). However, all the evidence presented so far is indirect in that it is inferred that the primary effect of the broadband noise (experiment 2), lowpass noise (experiment 3), and low-level stimuli (experiment 4) was to render distortion products inaudible. To explain the phase effects observed by Moore *et al.* (2006a) (and replicated here in experiment 1) at lower values of N, it must be assumed not only that distortion products were audible but that they were more audible (or had a greater effect on F0DLs) when the components were in COS phase than when they were in ALT phase. The results of Pressnitzer and Patterson (2001), using multitone complexes, suggest that this may indeed occur under certain conditions. They measured the distortion produced by harmonics 15–25 of a 100-Hz F0, with a level of 54 dB SPL per component. Using a beat-suppression technique, these authors estimated that a tone at the F0 had to be presented at between about 40 and 44 dB SPL to match the level of the distortion product when the complex was in COS phase, but when the complex was in ALT phase the level of the matching tone was found to be less than 35 dB SPL in one subject and was unmeasurable in the other subject. Although the level of their components was comparable to ours, the number of components and the N used were not, making direct predictions for our situation difficult. The stimuli of Buunen *et al.* (1974) were closer to those used in the present study. They measured distortion products for a three-component complex with an F0 of 200 Hz, centered at 2000 Hz, corresponding to our $N=9$ condition. Their stimuli were presented at a sensation level of 40 dB, and the lowest component was presented at a level 10 dB below that of the upper two components to maximize the relative level of the distortion products. They found clearly audible distortion products at the frequencies corresponding to one and two harmonics below N. Using a method of cancellation, they also found that the relative phase of the center component strongly affected the level of distortion products, as well as affecting the pitch strength of the F0. However, the relationship between the complex's phase and the level of the distortion products varied between the three subjects tested. Our final experiment attempted to estimate the relative levels of some distortion products in the presence of both the COS and ALT complexes. Our prediction was that the COS complex would produce higher-level distortion products than the ALT complex, thereby leading to a lower effective N in the COS than in the ALT case, even when the actual component frequencies were the same.

The technique used to provide an estimate of the relative levels of the distortion products involves presenting the three-tone complex along with an additional tone that was mistuned from the frequency of an expected distortion product by 4 Hz to produce an audible beating sensation with the distortion product, which should be maximal when the additional tone and the distortion product were at comparable effective levels. Listeners were asked to adjust the level of the additional tone to the level at which the beats were most prominent.

B. Methods

1. Subjects

Seven normal-hearing listeners participated in this experiment (ages ranged from 20 to 38). One of these listeners also took part in experiment 1.

2. Stimuli

Three-component complexes similar to those used in experiment 1 were used in this experiment. One difference was that the F0 of the complexes was fixed instead of roved across trials. The value of N was fixed at 8, and the F0 was 222.22 Hz, so that the frequency of the center harmonic was equal to 2000 Hz. The other stimulus difference with experiment 1 was that each complex lasted for 1150 ms, instead of 400 ms, to allow for multiple beat cycles. Each complex was accompanied by a "probe" pure tone with a frequency that was 4 Hz below the frequency of a potential distortion product. This probe tone was gated on 200 ms after the onset of the complex and off 200 ms before the offset of the complex. The total duration of the probe was therefore 750 ms (three cycles of the 4-Hz beat modulation), and it was gated on and off with 20-ms raised-cosine ramps. The distortion products tested were the three consecutive harmonic frequencies below the complex $(2f_1-f_2, 3f_1-2f_2, 4f_1-3f_2)$, as well as the F0 itself (the difference tone; f_2-f_1), where f_1 and f_2 denote the frequencies of the lowest and center harmonics in the complex, respectively. The 4-Hz deviation between the frequency of the probe tone and the frequency of the targeted distortion product was chosen based on informal pilot tests, which revealed that the depth of beats at 4 Hz was relatively easy to judge. No background noise was used in this experiment.

3. Procedure

Listeners were instructed to adjust the level of the probe tone until they found the level at which the 4-Hz beating sensation was most salient. The level of the probe tone could be adjusted after each stimulus presentation by three pairs of buttons on the computer screen. The upper, middle, and lower pairs of buttons adjusted the test-tone level by steps of 6, 3, and 1 dB, respectively. Listeners were instructed to try a wide range of levels on each run and to bracket the point of maximum saliency, first using largest step sizes and then the smaller ones. At the beginning of a match, the level of the probe tone was drawn randomly from a uniform discrete range including values between −25 and +25 dB around 50.2 dB SPL in 5-dB steps. The level of the probe tone was not allowed to exceed the limits of this 50-dB range. Each listener produced six matches in each condition.

C. Results and discussion

Beating was heard by subjects at all the test frequencies considered. The mean levels of the probe tones that were judged by subjects to produce the most salient beats are shown in Fig. 6. Overall, there was a small but significant trend for the level of the matching tone to be higher for the COS condition than for the ALT condition $[F(1,6)$

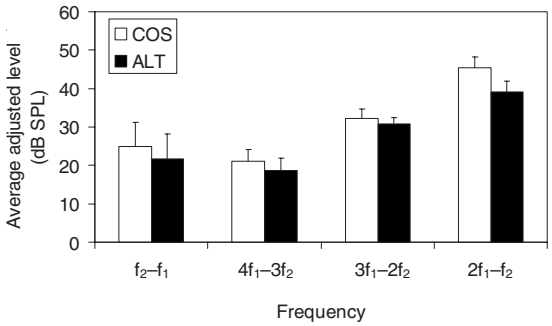


FIG. 6. Mean levels (in dB SPL) to which the matching probe tone was adjusted to produce the most salient beating sensation with the distortion products at different frequencies. Error bars represent ±1 standard error of the mean. See text for details of the different conditions.

$=96.67$, $p<0.0005$]. In the original experiment of Moore *et al.* (2006b), the background noise was set to 30 dB SPL/ERB_N, which would have resulted in masked thresholds of between 25 and 30 dB SPL. This suggests that only the $2f_1-f_2$ distortion product would have been reliably above masked threshold in the noise; no other average effective distortion product level fell above 33 dB SPL. Interestingly, the most audible ($2f_1-f_2$) distortion product also showed the largest mean difference between probe-tone levels in the COS and ALT conditions of about 6 dB (paired *t*-test: $t=4.29$, df=6, $p=0.005$).

The results suggest that at least the distortion product closest to the stimulus would have been audible in our experiment 1, which replicated that of Moore *et al.* (2006b). The higher level of the distortion product in the COS condition may have made it more salient and may therefore have lowered the effective value of *N* by 1 in the COS condition but not in the ALT condition. This in turn may have led to the effects of phase that were observed at intermediate values of *N* where the F0DLs were low.

VII. GENERAL DISCUSSION

Experiments 1 and 1a replicated the findings of Moore *et al.* (2006a) that F0DLs with three-tone harmonic complexes could be small, suggesting the use of TFS cues, and yet still be dependent on the component phase relationships, suggesting that the components were unresolved. However, experiments 2–4 showed that the results were different when distortion products were rendered inaudible, either by increasing the background noise level (experiment 2 and 3) or decreasing the overall stimulus level (experiment 4). Under these conditions, phase effects only emerged once F0DLs were high, implying the use of temporal envelope cues. The results from experiment 5 provided more direct evidence that the cubic distortion product just below the primary components would have been audible in the Moore *et al.* (2006a) experiment and may have been lower in amplitude in the ALT-phase than in the COS-phase conditions. The differences between the ALT and COS conditions found in their experiment (and our experiment 1) at intermediate values of *N* may have been due to the differential audibility of distortion rather than to differences in the TFS representations.

Overall, the results from the present study (experiments 2–4) are consistent with earlier studies using more harmonics (Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2006b; a) in showing that no phase effect is observed for conditions that produce small F0DLs. Based on these results, some of the conclusions from Moore *et al.* (2006a) require re-examination. As outlined in Sec. I, one rationale of the experiment of Moore *et al.* (2006a) was to distinguish between two interpretations of the earlier results of Moore and Moore (2003), involving the pitch of frequency-shifted harmonic complexes. Moore *et al.* (2006a) interpreted their finding of small F0DLs in conjunction with phase sensitivity as indicating that listeners had access to TFS within the waveform of unresolved harmonics. As phase sensitivity is not found when distortion products are rendered inaudible, the results of Moore *et al.* (2006a) can no longer be taken as evidence that listeners have access to TFS to code the F0 of unresolved harmonic tone complexes.

Another implication of the present results relates to studies of pitch perception and TFS processing in hearing-impaired listeners (e.g., Moore *et al.*, 2006b; Hopkins and Moore, 2007). These studies have shown that hearing-impaired listeners often do not exhibit low F0DLs for low or intermediate values of *N*. If, as suggested by Moore *et al.* (2006a), normal-hearing listeners can use TFS information from unresolved harmonics to code F0, poorer performance in hearing-impaired listeners might be interpreted as implying that they have a selective deficit in processing TFS. Similar claims have been made with regard to hearing-impaired listeners' ability to use TFS information in speech (e.g., Lorenzi *et al.*, 2006; Hopkins *et al.*, 2008). An alternative explanation is that the limited benefit of TFS information for hearing-impaired listeners relates to the well-known effect of broadened auditory filters, poorer frequency selectivity, and lower-amplitude (or absent) distortion products (Moore, 2007). For instance, Bernstein and Oxenham (2006b) found that the transition between good and poor F0DLs in hearing-impaired listeners corresponded well with other measures of frequency selectivity, suggesting that reduced frequency selectivity and fewer resolved harmonics may account for the poorer F0 discrimination associated with hearing impairment. Broader filters have been proposed as one reason for poorer TFS sensitivity in hearing-impaired listeners (Moore, 2008). However, given the lack of conclusive evidence that normal-hearing listeners use TFS to code F0 using unresolved harmonics, it may not be necessary to posit any deficits beyond frequency selectivity in order to explain the results found for hearing-impaired listeners in tests that have been designed so far to probe potential TFS deficits.

As mentioned in the discussion of experiment 2, the fact that the transition between good and poor F0DLs seems to occur when the average *N* is between 6 and 7 does not appear to be consistent with earlier studies, which found the transition to occur between *N* values of 9 and 12 (e.g., Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003). Whether this is due to the number of harmonics in each complex or to the value of F0 remains to be determined. In either case, it presents a challenge to the view that the transition between good and poor F0DLs is intimately linked

to auditory filter bandwidths. Such questions have been raised before, but primarily in conjunction with the lower limit of pitch perception (e.g., Krumbholz *et al.*, 2000). This question leads to a more general note of caution regarding the definition of resolved and unresolved components. The present study used the rationale forwarded by Moore *et al.* (2006a) to distinguish between resolved and unresolved harmonics, as well as between temporal envelope and temporal fine structure. However, this approach should not be interpreted as implying that the distinction between resolved and unresolved harmonics, or between temporal envelope and fine structure, is always clear cut (Shackleton and Carlyon, 1994). For instance, it is possible that situations exist in which stimulus components may be sufficiently resolved for one purpose (such as estimating the frequencies of individual components) and yet sufficiently unresolved for another purpose (such as perceiving interactions between neighboring components). One example of such a situation is the perception of beats between two tones a semitone apart, where the individual frequencies can be heard, but where the interactions (beats) are also audible. Similarly, stimuli that are not spectrally resolved (in terms of producing large peaks in a spectral excitation pattern) may nevertheless be resolved via a temporal code that involves phase-locking to the component frequency (e.g., Sachs *et al.*, 1983). In summary, it is unlikely that resolvability is a binary variable, and caution should be used in generalizing across different measures of resolvability.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (Grant No. R01 DC 05216). We thank Brian Moore, Daniel Pressnitzer, one anonymous reviewer, and the associate editor, Richard Freyman, for numerous helpful comments on previous versions of this paper.

[1]For the purposes of analysis, the two subjects who participated in both experiments were treated as independent subjects in each experiment. However, the conclusions remained the same if the two subjects were excluded from experiment 1a to maintain completely independent samples (and more equal sample sizes) in each experiment.

[2]The design of this analysis was complicated by the fact that some of the listeners took part in more than one experiment. As listeners who had lower thresholds on average (compared to other listeners) in one experiment also tended to have lower thresholds in another experiment, the data were not uncorrelated across experiments. In order to overcome this problem, we subtracted from each (log-transformed) threshold measured in a given listener the mean of the (log-transformed) thresholds measured in this same listener across all experiments and conditions in which the listener took part. This transformation left the effects of phase and harmonic number within each listener unchanged, but reduced across-experiment correlations in the data. Following this transformation, no significant main effect of experiment [$F(2,20)=3.32$, $p=0.057$] or interactions with experiment and other factors were observed [experiment$\times N$ interaction: $F(10.78,107.76)=1.03$, $p=0.423$; experiment$\times$phase interaction: $F(2,20)=0.03$, $p=0.970$; three-way interaction: $F(9.02,90.24)=0.439$, $p=0.911$]. These conclusions remained unchanged when the ANOVA was performed directly on the measured thresholds, without first subtracting the mean from each listener's data. Here and in the main text, non-integer degrees of freedom in the reported F statistics reflect the application of the Huynh–Feldt correction.

[3]These p values include Bonferroni's correction for multiple comparisons.

Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?" J. Acoust. Soc. Am. **113**, 3323–3334.

Bernstein, J. G., and Oxenham, A. J. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. **117**, 3816–3831.

Bernstein, J. G., and Oxenham, A. J. (**2006a**). "The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level," J. Acoust. Soc. Am. **120**, 3916–3928.

Bernstein, J. G., and Oxenham, A. J. (**2006b**). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," J. Acoust. Soc. Am. **120**, 3929–3945.

Bernstein, J. G., and Oxenham, A. J. (**2008**). "Harmonic segregation through mistuning can improve fundamental frequency discrimination," J. Acoust. Soc. Am. **124**, 1653–1667.

Buunen, T. J. F., Festen, J. M., Bilsen, F. A., and van den Brink, G. (**1974**). "Phase effects in a three-component signal," J. Acoust. Soc. Am. **55**, 297–303.

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?" J. Acoust. Soc. Am. **95**, 3541–3554.

de Boer, E. (**1956**). *On the "Residue" in Hearing* (University of Amsterdam, Amsterdam).

Fleischer, H. v. (**1976**). "Über die Wahrnehmbarkeit von Phasenänderungen (On the perception of phase changes)," Acustica **35**, 202–209.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Goldstein, J. L. (**1967**). "Auditory nonlinearity," J. Acoust. Soc. Am. **41**, 676–689.

Goldstein, J. L. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1516.

Hoekstra, A. (**1979**). *Frequency Discrimination and Frequency Analysis in Hearing* (Institute of Audiology, University Hospital, Groningen, The Netherlands).

Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

Houtsma, A. J. M., and Goldstein, J. L. (**1972**). "The central origin of the pitch of complex tones: Evidence from musical interval recognition," J. Acoust. Soc. Am. **51**, 520–529.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.

Kohlrausch, A., and Sander, A. (**1995**). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets," J. Acoust. Soc. Am. **97**, 1817–1829.

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1170–1180.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Licklider, J. C. R. (**1954**). "'Periodicity' pitch and 'place' pitch," J. Acoust. Soc. Am. **26**, 945.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Moore, B. C. J. (**1977**). "Effects of relative phase of the components on the pitch of three-component complex tones," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 349–358.

Moore, B. C. J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues* (Wiley, Chichester).

Moore, B. C. J. (**2008**). "The role of temporal fine structure in normal and impaired hearing," in *Auditory Signal Processing in Hearing-Impaired Listeners: First International Symposium on Auditory and Audiological Research (ISAAR 2007)*, edited by T. Dau, J. M. Buchholz, J. M. Harte,

and T. U. Christiansen (Centertryk A/S, Helsingor, Denmark), pp. 249–262.
Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (**2006a**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480–490.
Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (**2006b**). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure," Hear. Res. **222**, 16–27.
Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcantara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205–224.
Moore, G. A., and Moore, B. C. J. (**2003**). "Perception of the low pitch of frequency-shifted complexes," J. Acoust. Soc. Am. **113**, 977–985.
Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.
Oxenham, A. J., and Dau, T. (**2001**). "Towards a measure of auditory-filter phase response," J. Acoust. Soc. Am. **110**, 3169–3178.
Oxenham, A. J., and Keebler, M. V. (**2007**). "Pitch perception: Frequency selectivity and temporal coding," in Auditory Signal Processing in Hearing-Impaired Listeners (ISAAR 2007), edited by T. Dau, J. M. Buchholz, J. M. Harte, and T. U. Christiansen (Centrik A/S, Helsingor, Denmark), pp. 273–279.
Plack, C. J., Oxenham, A. J., Popper, A. N., and Fay, R., eds. (**2005**). *Pitch: Neural Coding and Perception* (Springer Verlag, New York.
Pressnitzer, D., and Patterson, R. D. (**2001**). "Distortion products and the pitch of harmonic complex tones," in *Physiological and Psychophysical Bases of Auditory Function*, edited by J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht), pp. 97–103.
Sachs, M. B., Voigt, H. F., and Young, E. D. (**1983**). "Auditory nerve representation of vowels in background noise," J. Neurophysiol. **50**, 27–45.
Schouten, J. F. (**1940**). "The residue and the mechanism of hearing," Proc. K. Ned. Akad. Wet. **43**, 991–999.
Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). "Pitch of the residue," J. Acoust. Soc. Am. **34**, 1418–1424.
Schroeder, M. R. (**1970**). "Synthesis of low peak-factor signals and binary sequences with low autocorrelation," IEEE Trans. Inf. Theory **16**, 85–89.
Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.
Shamma, S., 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.
Smoorenburg, G. F. (**1972a**). "Audibility region of combination tones," J. Acoust. Soc. Am. **52**, 603–614.
Smoorenburg, G. F. (**1972b**). "Combination tones and their origin," J. Acoust. Soc. Am. **52**, 615–632.
Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.
Wightman, F. L. (**1973**). "The pattern-transformation model of pitch," J. Acoust. Soc. Am. **54**, 407–416.

Release from informational masking in children: Effect of multiple signal bursts

Lori J. Leibold[a] and Angela Yarnell Bonino
Department of Allied Health Sciences, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

(Received 17 April 2008; revised 2 February 2009; accepted 4 February 2009)

This study examined the degree to which increasing the number of signal presentations provides children with a release from informational masking. Listeners were younger children (5–7 years), older children (8–10 years), and adults. Detection thresholds were measured for a sequence of repeating 50-ms bursts of a 1000-Hz pure-tone signal embedded in a sequence of 10- and 50-ms bursts of a random-frequency, two-tone masker. Masker bursts were played at an overall level of 60-dB sound pressure level in each interval of a two-interval, forced choice adaptive procedure. Performance was examined for conditions with two, four, five, and six signal bursts. Regardless of the number of signal bursts, thresholds for most children were higher than thresholds for most adults. Despite developmental effects in informational masking, however, masked threshold decreased with additional signal bursts by a similar amount for younger children, older children, and adults. The magnitude of masking release for both groups of children and for adults was inconsistent with absolute energy detection. Instead, increasing the number of signal bursts appears to aid children in the perceptual segregation of the fixed-frequency signal from the random-frequency masker as has been previously reported for adults [Kidd, G., *et al.* (2003). J. Acoust. Soc. Am. **114**, 2835–2845]. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3087435]

PACS number(s): 43.66.Lj, 43.66.Dc [RYL] Pages: 2200–2208

I. INTRODUCTION

A child's typical environment contains multiple sources of competing sounds. Each source generates acoustic waveforms consisting of individual frequency components that change over time. Many sound sources can be active at the same time and the child receives a mixture of overlapping waveforms from these multiple sources. In order to develop speech and language in this complex environment, the child must determine which frequency components were generated both sequentially and simultaneously from the same source. This process has been referred to in the literature as sound source segregation or sound source determination (e.g., Bregman, 1990; Yost, 1991).

Results from studies investigating children's speech perception in noise are consistent with the idea that the ability to perform sound source determination is immature during childhood. For example, children require a higher signal-to-noise ratio than adults to recognize speech in the presence of competing noise or speech maskers (e.g., Elliott *et al.*, 1979; Litovsky, 2005; Wightman and Kistler, 2005). Of particular interest to the study of sound source determination in children, these child-adult differences are pronounced for nonsensory or "perceptual" masking (Carhart *et al.*, 1969). Hall *et al.* (2002) observed larger developmental effects in speech recognition in the context of perceptual masking (i.e., spondee words presented in a two-talker masker) than in the context of "energetic" masking (i.e., words presented in speech-shaped noise). Note, however, that children's decreased performance on speech-in-noise tasks might also reflect their lack of experience using acoustic cues across different contexts. Moreover, the relation between sound source determination and auditory masking might be bidirectional for children. Increased masking might result in reduced sound source determination during childhood or vice versa.

Direct evidence supporting the hypothesis that infants and children have difficulty performing sound source determination is provided by studies of auditory stream segregation (Bregman, 1990). Auditory stream segregation refers to the ability to group incoming waveforms into separate auditory streams on the basis of acoustic cues that promote temporal coherence. For adults, these cues can include spatial separation, spectral separation, spectral profile, harmonicity, temporal onsets and offsets, and temporal modulations (e.g., Darwin and Carlyon, 1995; Yost, 1997). Results from several studies of auditory stream segregation during infancy indicate that this process is functional early in life (e.g., Winkler *et al.*, 2003) and that infants use many of the same acoustic cues that adults use to segregate auditory streams (Demany, 1982; Fassbender, 1993; McAdams and Bertoncini, 1997). Larger acoustic separations have been used for testing infants than for testing adults, however, and procedural modifications introduced to test infants limit direct comparisons with adult data. Thus, while the process of auditory stream segregation appears functional early in life, it is not clear how accurately this process operates as the child enters the preschool and school-aged years.

Recent data reported by Sussman *et al.* (2007) indicate that the ability to perform auditory stream segregation continues to develop well into childhood. School-aged children (5–11 years) and adults were presented with a sequence of

[a]Author to whom correspondence should be addressed. Electronic mail: leibold@med.unc.edu

 0001-4966/2009/125(4)/2200/9/$25.00

alternating pure tones. The difference in frequency of the alternating tones was varied and listeners were asked to report whether they heard a single integrated auditory stream or two distinct streams. Compared to adults, children required a greater frequency difference between alternating tones before indicating they heard two separate streams. Sussman *et al.* (2007) suggested that the process of auditory stream segregation requires maturation and perceptual experience to fully develop.

We are unaware of other studies describing the efficiency and accuracy of sound source determination during childhood. Thus, it is not clear whether children's abilities to use cues in addition to frequency proximity also follow a prolonged course of development. This lack of information reflects, in part, how challenging it can be for young children to perform the conventional auditory streaming tasks. Interpreting data obtained from these studies is further complicated by the potential for differences in response bias across children and adults using conventional auditory streaming procedures. Differences in response bias can occur as a result of criterion placement and may be responsible for individual differences in the streaming percept across adult listeners for intermediate or ambiguous frequency separations (e.g., Rose and Moore, 1997). Furthermore, studies of complex auditory perception using single-interval tasks have reported developmental changes in response bias (e.g., Leibold and Werner, 2006). Thus, interpreting child-adult differences in auditory streaming is challenging. Observed developmental changes might reflect immaturity in auditory streaming mechanisms. Alternatively, these age effects might reflect differences in the underlying decision processes used by children and adults.

An alternative approach for examining the development of sound source determination is to examine whether children benefit from stimulus manipulations thought to aid in sound source determination by providing a release from "informational" masking (e.g., Watson *et al.*, 1975; Neff and Green, 1987; Kidd *et al.*, 1994). In this context, informational masking refers to masking produced by energy remote in frequency from the signal, that is, masking produced despite an adequate sensory-neural representation of the signal. For simultaneous masking of a pure-tone signal, these effects are largest when stimulus uncertainty is created by introducing random variations in the spectral content of a multi-tonal masker. For example, adults' thresholds for a fixed-frequency, pure-tone signal can be elevated by as much as 50 dB when the spectral content of a multi-tonal masker is varied on each presentation (e.g., Neff and Green, 1987). One advantage of using this approach is that measures of masking release can be obtained efficiently by estimating thresholds using multiple-interval, forced-choice adaptive procedures not likely to be influenced by the placement of the listener's response criterion (e.g., Marshall and Jesteadt, 1986).

The mechanisms responsible for informational masking are not fully understood, but a failure of sound source determination appears to be responsible for a substantial portion of the masking observed in these conditions (e.g., Kidd *et al.*, 1994, 2002; Neff, 1995; Durlach *et al.*, 2003). Evidence supporting this hypothesis is provided by studies that have manipulated stimulus properties thought to promote sound source determination, including spatial separation, asynchronous temporal onsets, and dissimilar temporal modulations (e.g., Darwin and Carlyon, 1995; Yost, 1997). A substantial release from informational masking is observed for most adult listeners when these cues are introduced (e.g., Kidd *et al.*, 1994; Neff, 1995; Oh and Lutfi, 1998; Arbogast *et al.*, 2002; Durlach *et al.*, 2003).

On average, children appear to be more susceptible to informational masking than adults (e.g., Allen and Wightman, 1995; Oh *et al.*, 2001; Wightman *et al.*, 2003). For example, Oh *et al.* (2001) found that the average threshold for detection of a 1000-Hz tone embedded in a random-frequency, multi-tonal masker was about 20 dB higher for children (4–16 years) than for adults. Given children's increased susceptibility to informational masking relative to adults and the mounting evidence that informational masking reflects difficulties performing sound source determination, identifying stimulus cues that benefit children's performance on these tasks has the potential to provide valuable information regarding the development of sound source determination. This approach has been used to examine whether informational masking in children can be reduced by presenting the signal and the masker to different ears (Wightman *et al.*, 2003; Hall *et al.*, 2005), asynchronous temporal onsets (Hall *et al.*, 2005; Leibold and Neff, 2007), spectro-temporal coherence (Hall *et al.*, 2005), spatial separation (Litovsky, 2005), and auditory-visual integration (Wightman *et al.*, 2006). As Hall *et al.* (2005) discussed, results from these studies suggest that the extent to which children benefit from stimulus cues that promote sound source determination for adults depends on the specific cue manipulated. For example, a smaller release from informational masking has been observed for children than for adults when the target signal and the masker are presented dichotically (Wightman *et al.*, 2003; Hall *et al.*, 2005). Both Wightman *et al.* (2003) and Hall *et al.* (2005) found that children receive limited benefit when a fixed-frequency pure-tone signal and a random-frequency multi-tonal masker are presented to opposite ears, a manipulation associated with a large masking release for adults. In contrast, informational masking can be effectively reduced for most children when the onset of a pure-tone signal is delayed relative to the onset of a multi-tonal masker. Children and adults receive a similar average masking release for signal-masker onset asynchronies of 120 ms (Hall *et al.*, 2005) or 100 ms (Leibold and Neff, 2007). Thus, whereas most children demonstrate substantial benefit for a temporal onset asynchrony cue, they demonstrate reduced efficiency for other stimulus cues such as contralateral presentation of the signal and masker. Discrepancies across studies indicate that the relative benefit or salience of stimulus cues that promote a release from informational masking changes with development. However, data remain limited and the benefit provided by several cues shown to aid adults' performance has not yet been determined for children. In addition, most developmental studies have examined performance across conditions with relatively large acoustic differences. Few developmental studies have parametrically varied

the strength of the grouping cue or examined the extent to which children benefit from smaller manipulations.

This study examined the degree to which children benefit from a potential spectro-temporal grouping cue, multiple presentations of a constant-frequency, pure-tone signal. A modified "multiple-burst different" informational masking paradigm (Kidd *et al.*, 1994) was used to parametrically examine child-adult differences in informational masking release for conditions with two, four, five, and six bursts of a 50-ms, 1000-Hz tone embedded in a sequence of random-frequency, two-tone masker bursts. Multiple presentations of the fixed-frequency signal bursts appear to contrast with the random-frequency masker bursts to form a coherent auditory stream and reduce masking for adults (Kidd *et al.*, 2003; Huang and Richards, 2006). Children, however, appear to listen less selectively than adults in complex listening conditions (e.g., Lutfi *et al.*, 2003) and may require more signal presentations than adults to form a coherent stream and perceptually segregate the signal from the masker.

II. METHODS

A. Listeners

Fifteen children (5–10 years) and eight adults (20–29 years) participated in all conditions. Two groups of children were studied: (1) seven younger children aged 5–7 years and (2) eight older children aged 8–10 years. Younger children had a mean age (years:months) of 6:6 (range=5:5–7:6), older children had a mean age of 9:5 (range=8:6–10:5), and adults had a mean age of 24:1 (range=20:0–29:5). Listeners were required to pass a hearing screening prior to testing (re: ANSI, 1996) and had no known history of chronic ear disease. Testing occurred in a single-walled, sound-treated room. Adults and older children completed testing in a 2-h visit. Younger children were typically tested in two 1-h visits. Regular breaks were provided during testing, after the completion of two to three conditions for children and between blocks for adults. Three additional children were tested, but were excluded from data analysis. One child (age 5) did not meet the training criteria, discussed below. A second child (age 5) adapted out of bounds for one condition and did not provide sufficient data to estimate a threshold. Data for a third child (age 7) were excluded because of unusually variable threshold estimates across test sessions. A closer examination of this child's adaptive tracks revealed inconsistent responses and high variability (>5 dB) of reversals.

B. Stimuli and conditions

Following the multiple-burst paradigm of Kidd *et al.* (1994), the masker was a sequence of 10- and 50-ms tone bursts (5-ms, $\cos^2$, rise/fall ramps). There was no temporal overlap between successive masker bursts, resulting in a total duration of 500 ms for each masker sample. Individual masker bursts within the sequence were comprised of two frequency components drawn randomly from a uniform distribution with a range of 300–3000 Hz, excluding 920–1080 Hz. The frequency range from 920 to 1080 Hz extends beyond the equivalent rectangular bandwidth centered on 1000 Hz (Glasberg and Moore, 1990) and was excluded to reduce energy-based masking. The overall level of each burst was 60-dB sound pressure level (SPL) (57 dB/component).

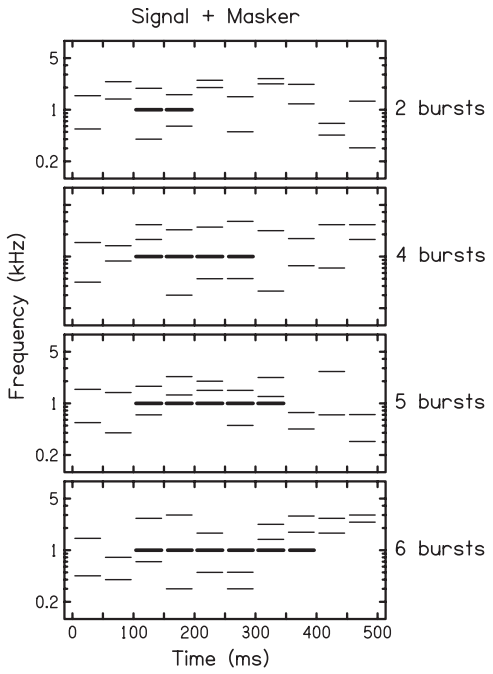


FIG. 1. Schematic representations of the stimuli and conditions are shown as spectrograms. The signal interval of a 2IFC trial with two, four, five, or six bursts of a 1000-Hz tone (bold) are illustrated. For all conditions, the masker consisted of ten bursts of two-component random-frequency maskers (light shading).

The signal was a sequence of 50-ms bursts of a 1000-Hz tone (5-ms, $\cos^2$, rise/fall ramps). Across conditions, the number of signal bursts was varied. Performance for conditions with two, four, five, and six bursts was examined in the presence of the sequential masker. The choice of conditions was based on data reported by Kidd *et al.* (2003) for trained adult listeners using similar conditions as well as pilot data collected from untrained adult and child listeners using the current stimuli. Performance was not assessed for a one-burst condition in the present study. This condition was not tested because both the data reported by Kidd *et al.* (2003) and the pilot data indicated that most children and some adults had difficulty detecting a single, 50-ms burst of the 1000-Hz tone presented in a random, multiple-burst masker. The signal bursts were turned on and off synchronously with the masker bursts, but the onset of the initial signal burst was delayed by 100 ms relative to the first masker burst. Thus, the onset of the first signal burst coincided with the onset of the third masker burst. In addition, a minimum of two masker bursts were played following the offset of the signal bursts. The expectation was that the 100-ms asynchronous onset and offset between the signal and masker bursts would aid in the perceptual segregation of the signal from the masker to a similar degree for both adults and children (Hall *et al.*, 2005; Leibold and Neff, 2007). Schematic representations of the stimuli for the signal-plus-masker interval of one two-interval, forced choice (2IFC) trial are provided in Fig. 1. Thresholds were also measured for the two-burst and six-burst signals in quiet.

Stimuli were played through a 24-bit digital-to-analog converter (Digital Audio Laboratories) at a sampling rate of 20 kHz. Stimuli were presented monaurally to the listener's left ear via Sennheiser HD-25 earphones. Twenty-five masker sequences were generated and stored. The computer randomly selected a different masker sample from this file on each presentation. The same 25 masker samples were used for all masker conditions.

C. Procedure

Prior to testing, listeners completed training for the two-burst signal presented in quiet and both the two-burst and six-burst signals embedded in the random masker sequence. Children sat in front of a computer monitor. Using a two-interval, forced-choice procedure, each trial consisted of two successive observation intervals (indicated by "1" and "2" on the computer monitor). The interstimulus interval was 500 ms. The signal was presented at a clearly audible level during training, based on pilot data. The initial training level was 80-dB SPL. The training level was increased to 90-dB SPL for children who had difficulty performing the task at the initial level. The signal occurred in either interval with equal *a priori* probability and children were instructed to verbally indicate which interval contained the signal. An experimenter sat inside the booth with the child, initiated trials, and entered responses. Correct responses were rewarded by an image presented on the monitor in a video game format. Adults were tested using the same procedure, but were alone in the booth, self-initiated trials, and entered responses directly. Training continued until the listener correctly responded to a minimum of four correct responses within five consecutive trials. As noted previously, one child (age 5) was unable to meet the training criteria.

Following training, thresholds for the 1000-Hz signal were measured with the two-interval, forced-choice procedure used for training, but combined with a two-down, one-up adaptive paradigm that estimated 70.7% correct on the psychometric function (Levitt, 1971). For the first testing block, the individual training level was used as the starting level of the adaptive track. For the second and subsequent testing blocks, the starting level was 10–15 dB above the initial threshold estimate, adjusted for individual listeners. The initial step size was 4 dB, followed by a step size of 2 dB starting with the third reversal. The maximum allowable signal level was 96-dB SPL. For masked conditions, the adaptive track was terminated after 12 reversals and threshold was determined by averaging the levels for the last 10 reversals. For quiet conditions, the adaptive track was terminated after eight reversals and threshold was determined by averaging the levels for the last six reversals.

The order of testing was randomized across all conditions within blocks for each listener. All listeners completed two blocks of testing. Additional blocks were completed if the first two threshold estimates varied by more than 5 dB. For younger children, additional blocks were run for 36% of quiet conditions and 54% of random two-tone conditions. For older children, additional blocks were run for 19% of quiet conditions and 50% of random two-tone conditions. For adults, additional blocks were run for 6% of quiet conditions and 44% of random two-tone conditions. The two threshold estimates with the best agreement were used to determine a listener's average threshold for each condition.

III. RESULTS

Individual threshold estimates for the two-burst and six-burst signals presented in quiet are provided in Table I. Figure 2 shows average quiet thresholds for younger children (open triangles), older children (open squares), and adults (filled circles) as a function of the number of signal bursts (2 versus 6). Error bars represent ± 1 standard error (SE) of the mean threshold across listeners within each age group. For both groups of children, the average threshold for the six-burst signal in quiet is about 2 dB lower than the average threshold for the two-burst signal in quiet. The average adult threshold for the six-burst signal in quiet is about 4 dB lower than the average adult threshold for the two-burst signal in quiet.

A repeated-measures analysis of variance (ANOVA) of threshold was consistent with the trends observed in Fig. 2. No significant differences in absolute threshold across the three age groups were observed [$F(2,20)=2.1$; $p=0.1$], indicating similar performance in quiet for younger children, older children, and adults. However, this analysis showed a significant effect of number of signal bursts [$F(1,20)=9.1$; $p=0.007$], indicating thresholds in quiet decreased as the number of signal bursts increased from 2 to 6. There was no interaction between age group and number of signal bursts ($p=0.7$). Thus, increasing the number of signal bursts in quiet did not affect adults' performance to a significantly greater degree than it affected children's performance.

Individual thresholds for bursts of the 1000-Hz signal presented in the multiple-burst masker are also provided in Table I. Figure 3 shows average masked threshold for younger children (open triangles), older children (open squares), and adults (filled circles) as function of the number of signal bursts. Error bars represent ± 1 SE of the mean threshold across listeners within each age group. For all conditions, average masked threshold for both groups of children was higher than the average masked threshold for adults. Moreover, average masked thresholds for younger children were consistently higher than average masked thresholds for older children.

Despite age differences in susceptibility to masking, masked threshold decreased as the number of signal bursts increased for all three age groups. The largest improvement in average masked threshold was observed as the number of signal bursts increased from 2 to 4. The average threshold improvement in the four-burst condition relative to the two-burst condition was 15.6 dB (SE=7.3) for younger children, 14.2 dB (SE=3.5) for older children, and 9.5 dB (SE=1.3) for adults. A smaller, but progressive, improvement in threshold was observed as the number of signal bursts increased from 4 to 6. The average threshold improvement in the six-burst condition relative to the four-burst condition was 8.7 dB (SE=1.7) for younger children, 7.8 dB (SE=2.7) for older children, and 5.3 dB (SE=2.7) for adults.

TABLE I. Thresholds in decibel SPL for younger children (5–7 years), older children (8–10 years), and adults for the signal bursts presented in quiet and presented in the random masker sequence. The age of each listener is given in years:months.

	Quiet		Masker				
	2	6	2	4	5	6	Age
			Younger children				
YC1	12.5	26.3	90.6	83.4	72.0	73.9	5:5
YC3	8.4	3.1	76.8	52.9	57.4	40.0	5:10
YC3	13.4	8.4	87.2	82.6	90.7	76.4	6:1
YC4	8.8	4.8	72.9	70.9	62.6	64.4	6:11
YC5	4.8	2.7	90.0	82.1	71.5	68.6	7:2
YC6	15.3	7.3	89.9	82.6	72.0	71.1	7:4
YC7	10.0	7.1	83.9	27.7	28.8	26.9	7:6
Mean	10.5	8.5	84.5	68.9	65.0	60.2	
SE	1.3	3.1	2.7	8.0	7.2	7.2	
			Older children				
OC1	7.4	5.1	35.8	18.7	18.5	22.8	8:6
OC2	3.5	1.9	74.3	54.4	43.6	40.5	8:9
OC3	10.6	8.8	78.5	73.8	70.4	68.7	9:6
OC4	8.6	1.7	73.6	65.8	53.7	51.6	9:9
OC5	1.2	0.6	45.7	22.6	15.4	15.7	10:2
OC6	10.8	9.6	73.3	67.3	66.3	69.0	10:4
OC7	2.7	0.3	75.2	44.9	35.3	30.1	10:5
OC8	6.0	3.3	75.1	70.7	78.1	57.6	10:5
Mean	6.3	3.9	66.4	52.3	47.7	44.5	
SE	1.3	1.3	5.7	7.7	8.3	7.2	
			Adults				
A1	4.5	2.8	25.7	17.6	16.8	18.0	20
A2	2.5	−2.3	37.6	21.2	15.6	10.0	20
A3	16.6	14.6	34.8	27.4	27.6	25.1	20
A4	2.5	0.7	76.2	64.7	56.1	60.7	21
A5	10.3	3.4	66.0	54.0	39.0	32.8	26
A6	9.6	5.1	30.4	22.2	20.1	16.9	27
A7	7.3	1.0	62.7	55.1	60.3	58.7	27
A8	6.5	4.5	70.3	65.2	60.3	62.7	29
Mean	7.5	3.7	50.5	40.9	37.0	35.6	
SE	1.7	1.8	7.2	7.3	6.9	7.7	

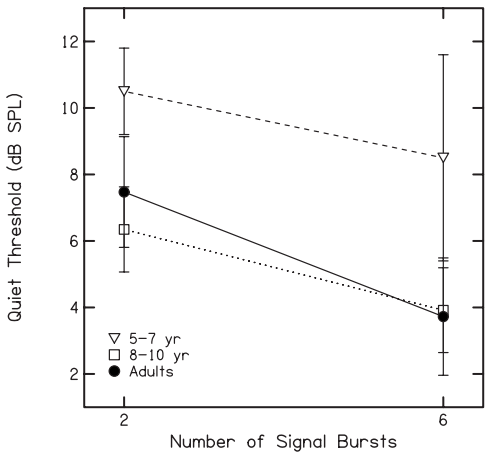


FIG. 2. Average quiet thresholds across listeners (with SEs) for each of the three age groups (open triangles for younger children, open squares for older children, and filled circles for adults) are presented for the two-burst and six-burst signals.

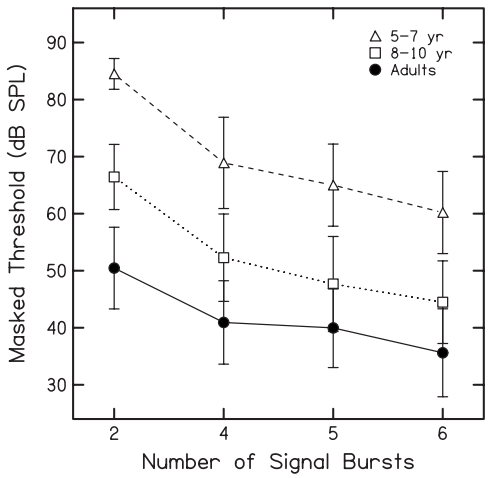


FIG. 3. Average masked thresholds across listeners (with SEs) for each of the three age groups (open triangles for younger children, open squares for older children, and filled circles for adults) are presented for the two-, four-, five-, and six-burst signals.

A repeated-measures ANOVA of threshold was performed with one between-subjects factor (age) and one within-subjects factor (number of signal bursts) to assess developmental effects in the amount of masking release associated with increasing the number of signal bursts. This analysis revealed a significant main effect of number of bursts [$F(3,60)=32.9$; $p<0.001$], indicating thresholds decreased as the number of bursts of the 1000-Hz signal increased. *Post hoc* pairwise comparisons (Bonferroni, using a criterion of $p<0.05$) indicated that thresholds were significantly higher in the two-burst condition than in the four-, five-, and six-burst conditions. Thresholds were also higher in the four-burst condition than in the six-burst condition. Thresholds were not significantly different across the four-burst and five-burst conditions ($p=0.06$) or across the five-burst and six-burst conditions ($p=0.2$). The main effect of age was also significant [$F(2,20)=4,4$; $p=0.03$], indicating developmental effects in susceptibility to masking for these conditions. *Post hoc* pairwise comparisons (Bonferroni, using a criterion of $p<0.05$) indicated that thresholds were significantly higher for younger children than for adults, but were not significantly different across older children and adults ($p=0.7$) or across older and younger children ($p=0.3$). No interaction between number of bursts and age was observed [$F(6,60)=0.6$; $p=0.7$], indicating that threshold decreased with increased signal bursts in a similar way for younger children, older children, and adults.

Considerable between-subjects variability in both amount of masking and masking release was observed for all three age groups, as shown by the error bars in Fig. 3 and the individual thresholds provided in Table I. Despite large individual differences between and within age groups, however, thresholds for all listeners decreased as the number of signal bursts increased. With the exception of one younger child, all listeners showed a masking release of 4 dB or greater as the number of signal bursts was increased from 2 to 4. The remaining listener (age 6:11) showed a masking release of 2 dB. All listeners, including the younger child (6:11) who showed a 2-dB improvement in threshold when the number of signal bursts increased from 2 to 4, showed a masking release as the number of signal bursts was increased from 2 to 6. Also note that the majority of listeners showed additional masking release as the number of signal bursts was increased from four to six bursts.

IV. DISCUSSION

A. Child-adult differences in susceptibility to informational masking

1. Group differences

The current results support the hypothesis that informational masking is greater for children than for adults. Thresholds for the detection of bursts of a 1000-Hz signal in the presence of a random-frequency masker sequence were elevated for all three age groups. However, younger children had significantly higher masked thresholds than adults. These results are in agreement with previous developmental studies of informational masking that have employed sequential multi-tonal maskers (Hall *et al.*, 2005), simultaneous multi-tonal maskers (e.g., Oh *et al.*, 2001; Wightman *et al.*, 2003), and speech maskers (e.g., Wightman and Kistler, 2005). The mechanisms responsible for these age-related changes in susceptibility to informational masking are not fully understood, but are not likely the consequence of immature sensory processing [reviewed by Werner (1996)]. Instead, children's increased susceptibility to informational masking appears to reflect developmental changes in nonsensory processes. For example, Lutfi *et al.* (2003) modeled children's responses in a pure-tone informational masking task, and their elevated thresholds suggest a reduced ability to attend selectively to the auditory filter containing the signal.

2. Individual differences

As in previous investigations of informational masking involving both children (e.g., Oh *et al.*, 2001) and adults (e.g., Neff and Dethlefs, 1995), a wide range of performance was observed within and across age groups. However, systematic differences in threshold variability were not observed across the three age groups. For example, the range of masked thresholds for the six-burst condition spanned a range of 47 dB for younger children and 53 dB for both older children and adults. In contrast, Hall *et al.* (2005) observed a smaller range of thresholds for adults (30 dB) than for children aged 4–9 years (50 dB) for an eight-burst signal embedded in a sequence of two-tone masker bursts with random spectra. Differences in methodology across studies make it difficult to determine the basis for this discrepancy. Note, however, that differences in threshold variability between the current data and Hall *et al.* (2005) reflect an increased range of threshold estimates for adults in the current study. A comparison of threshold estimates for children across the two studies indicates a similar range of performance.

B. Child-adult similarity in release from informational masking

The main purpose of this study was to determine whether there are developmental effects in informational masking release as the number of signal presentations increased. The results indicate this cue reduced masking for younger children (5–7 years), older children (8–10 years), and adults. Even with adults' lower thresholds in the baseline (two-burst) condition compared to younger children, the three age groups experienced similar release from masking as the number of signal bursts increased.

The pattern of results for untrained adult listeners in the current study is in general agreement with the pattern observed by Kidd *et al.* (2003) for trained adults. This agreement suggests that increasing the number of presentations of the signal, as opposed to the masker, is likely responsible for reducing informational masking using the multiple-burst different paradigm. Kidd *et al.* (2003) examined listeners' ability to detect a sequence of 60-ms, 1000-Hz signal bursts embedded in a masker consisting of 60-ms bursts of a random-frequency, eight-tone complex. In contrast to the current study in which the number of masker bursts was 10 for all conditions, Kidd *et al.* (2003) varied the number of masker bursts across conditions to match the number of sig-

nal bursts. In conditions with one, two, four, and eight bursts and an inter-burst interval of 0 ms, all listeners showed a masking release as the number of signal and masker bursts was increased. The current results corroborate these previous findings by confirming the effectiveness of providing listeners with a coherent signal across time for reducing informational masking (e.g., Kidd *et al.*, 2003; Huang and Richards, 2006). Inconsistent with the result of Kidd *et al.* (2003), however, the present study showed smaller reductions in threshold as the number of signal bursts increased. Whereas Kidd *et al.* (2003) found an average masking release of 24 dB as the number of masker bursts was increased from 2 to 4, an average masking release of 10 dB was observed here. Differences in stimuli are the most likely reason for this discrepancy. In addition to varying the number of both signal and masker bursts, Kidd *et al.* (2003) used masker bursts comprised of eight pure-tone components. Maskers with fewer components, such as the two-component maskers used in the current study, have been shown to produce less informational masking (e.g., Neff and Green, 1987) and thus, a reduced potential for masking release.

Increasing the number of presentations of a 1000-Hz signal was clearly an effective cue for children as well as adults. Despite considerable individual differences, children as young as 5 years of age were able to take advantage of the increased number of signal bursts. Though a direct comparison is not evident in the literature, Hall *et al.* (2005) also found that most children (5–10 years) benefited from stimulus cues believed to be related to spectro-temporal coherence. Inconsistent with the current results, however, Hall *et al.* (2005) observed a smaller average masking release for children than for adults. Perhaps the most important difference across the two studies is that Hall *et al.* (2005) assessed informational masking release related to manipulations made to the spectral properties of the masker. The current study examined informational masking release related to changes made to the temporal properties of the signal.

C. Relation between susceptibility to and release from informational masking

The finding that children's thresholds were generally higher than adults in the baseline (two-burst) condition complicates the interpretation of age differences in masking release. Increased susceptibility to informational masking in some children may have permitted larger reductions in threshold. Several observations are inconsistent with this explanation. First, children with the greatest release from masking did not have uniformly high thresholds in the baseline (two-burst) condition. In fact, the child (8:6) with the lowest threshold for the two-burst condition among all children showed a larger-than-average masking release (17 dB) as the number of signal bursts increased to four bursts. Second, regression analyses indicated that thresholds in the two-burst condition were not a significant predictor of the amount of masking release observed as the number of signal bursts was increased from two to four bursts for any age group (all p's >0.8). Third, a comparison of informational masking release among adults found no evidence of differences in masking release for the four adults with the lowest two-burst thresholds ($<$40-dB SPL) versus the four adults with the highest two-burst thresholds ($>$60-dB SPL). The average masking release for the four adults with the lowest thresholds was 10 dB. Similarly, the average masking release for the four adults with the highest thresholds was 9 dB. Together, these observations are inconsistent with the viewpoint that developmental effects in susceptibility to informational masking are responsible for the amount of benefit provided by increasing the number of signal presentations.

The average child-adult difference in threshold for the six-burst signal embedded in the random masker was 26 dB for younger children and 9 dB for older children. Thus, children's thresholds remain elevated compared to adults despite their ability to benefit from the increased number of signal bursts. The basis for children's increased susceptibility to informational masking relative to adults in the presence of an effective cue is unknown. However, this finding is consistent with previous studies that have examined developmental effects in susceptibility to and release from informational masking. For example, Leibold and Neff (2007) observed a similar effect of reducing masker-spectral variability and adding an onset/offset asynchrony cue for children and adults in the context of a simultaneous informational masking task. Nonetheless, Leibold and Neff (2007) reported that children's thresholds remained elevated compared to adults for conditions in which both of these grouping cues were available.

D. Energy detection versus perceptual segregation

One explanation for the observed masking release is that listeners combined or integrated stimulus energy across time to improve performance. It has been well documented that detection thresholds for a pure-tone signal presented in quiet or in broadband noise decrease as signal duration increases from about 10 to 200–300 ms [reviewed by Gerken *et al.* (1990)]. This phenomenon is typically referred to as temporal integration. Several models have been proposed to account for temporal integration, including models that assume that listeners integrate stimulus energy across 200–300 ms (e.g., Swets *et al.*, 1959; McKinley and Weber, 1994) and more recent models that assume that the listener's decision is based on information combined across multiple shorter "looks" or samples (Viemeister and Wakefield, 1991).

The current data for the signal presented in the random two-tone masker are inconsistent with energy-detector and multiple-looks models. For example, a decrease in threshold of approximately 10 dB/decade increase in signal duration is predicted using a simple energy-detector model that assumes linear integration. For the current stimuli, this prediction corresponds to an improvement in threshold of approximately 3 dB for the four-burst signal (200 ms) compared to the two-burst signal (100 ms). In contrast to the predicted improvement, the average observed masking releases were 16, 14, and 10 dB for younger children, older children, and adults, respectively. The discrepancy between predicted and observed performances for the random two-tone masker suggests that listeners are not basing their decisions on absolute energy detection. In addition, the average observed threshold

improvement in quiet across the six-burst and two-burst signals was about 2 dB for both groups of children and 4 dB for adults. Thus, whereas the observed threshold improvement was equal to (adults) or slightly less than (children) predicted for optimal energy detection in quiet, the magnitude of threshold improvement for all three age groups with increasing signal bursts was considerably greater than predicted in the presence of the random two-tone masker.

Previous work by Kidd *et al.* (2003) examined whether a multiple-looks model (Viemeister and Wakefield, 1991) could account for adults' performance on multiple-burst different conditions. Both the number of signal bursts and the silent interval between successive bursts (interburst interval) were manipulated. Consistent with the multiple-looks model, masking decreased as the number of bursts increased. However, several observations were inconsistent with the multiple-looks model. First, the observed magnitude of masking release exceeded the amount predicted by the model. Second, masking increased as the interburst interval increased. Third, the group-mean slope of the psychometric function increased as the number of signal bursts increased or as the interburst interval decreased. A similar increase in the slope of the psychometric function with reductions in informational masking has been observed by researchers using a simultaneous multi-tonal paradigm (e.g., Durlach *et al.*, 2005). Based on these observations, Kidd *et al.* (2003) suggested that increasing the number of bursts of the fixed-frequency signal contrasted with the random-frequency masker to provide an effective perceptual grouping cue. The magnitude of masking release and the discrepancy across quiet and masked conditions observed in the current study suggest that increasing the number of signal bursts strengthens perceptual coherence and reduces thresholds for children as well as for adults.

V. CONCLUSIONS

The results of the current study indicate that multiple presentations of a constant-frequency signal embedded in a random-frequency masker provide both children and adults with a robust cue for reducing informational masking. Although younger children were more susceptible to informational masking than adults, a similar decline in thresholds was observed as the number of signal bursts was increased for all age groups. Determining the relative importance of stimulus cues that reduce informational masking has the potential to provide insight into the mechanisms responsible for children's increased susceptibility to auditory masking. Further studies are needed to determine which cues are the most salient during development, how children integrate multiple cues to improve performance in the presence of competing sounds, and whether infants and children can benefit from similar acoustic cues in more natural environments.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD Grant No. R03 DC008389. We thank Caitlin Rawn and Laura Fleenor for assistance during data collection. We are grateful to Joseph Hall III and Emily Buss for helpful discussions and valuable comments.

Allen, P. A., and Wightman, F. (**1995**). "Effects of signal and masker uncertainty on children's detection," J. Speech Hear. Res. **38**, 503–511.
ANSI (**1996**). *Specification for Audiometers* (American National Standards Institute, New York), ANSI S3.6-1996.
Arbogast, T. L., Mason, C. R., and Kidd, G. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.
Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).
Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.
Darwin, C. J., and Carlyon, R. P. (**1995**). in *Hearing*, edited by B. C. J. Moore (Academic, London), pp. 387–424.
Demany, L. (**1982**). "Auditory stream segregation in infancy," Infant Behav. Dev. **5**, 261–276.
Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B. G., Colburn, H. S., and Kidd, G., Jr. (**2005**). "Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers," J. Acoust. Soc. Am. **118**, 2482–2497.
Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (**2003**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.
Elliott, L. L., Connors, S., Kille, E., Levin, S., Ball, K., and Katz, D. (**1979**). "Children's understanding of monosyllabic nouns in quiet and noise," J. Acoust. Soc. Am. **66**, 12–21.
Fassbender, C. (**1993**). *Auditory Grouping and Segregation Processes in Infancy* (Kaste, Norderstedt, Germany).
Gerken, G. M., Bhat, V. K. H., and Hutchinson-Clutter, M. H. (**1990**). "Auditory temporal integration and the power-function model," J. Acoust. Soc. Am. **88**, 767–778.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Hall, J. W. III, Buss, E., and Grose, J. H. (**2005**). "Informational masking release in children and adults," J. Acoust. Soc. Am. **118**, 1605–1613.
Hall, J. W. III, Grose, J. H., Buss, E., and Dev, M. B. (**2002**). "Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children," Ear Hear. **23**, 159–165.
Huang, R., and Richards, V. M. (**2006**). "Coherence detection: Effects of frequency, frequency uncertainty, and onset/offset delays," J. Acoust. Soc. Am. **119**, 2298–2304.
Kidd, G., Jr., Mason, C. R., and Arbogast, T. L. (**2002**). "Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **111**, 1367–1376.
Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.
Kidd, G., Mason, C. R., and Richards, V. M. (**2003**). "Multiple bursts, multiple looks, and stream coherence in the release from informational masking," J. Acoust. Soc. Am. **114**, 2835–2845.
Leibold, L. J., and Neff, D. L. (**2007**). "Effects of masker-spectral variability and masker fringes in children and adults," J. Acoust. Soc. Am. **121**, 3666–3676.
Leibold, L. J., and Werner, L. A. (**2006**). "Effect of masker-frequency variability on the detection performance of infants and adults," J. Acoust. Soc. Am. **119**, 3960–3970.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Litovsky, R. Y. (**2005**). "Speech intelligibility and spatial release from masking in young children," J. Acoust. Soc. Am. **117**, 3091–3099.
Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003**). "One factor underlies individual differences in auditory informational masking within and across age groups," Percept. Psychophys. **65**, 396–406.
Marshall, L., and Jesteadt, W. (**1986**). "Comparison of pure-tone audibility thresholds obtained with audiological and two-interval forced-choice procedures," J. Speech Hear. Res. **29**, 82–91.
McAdams, S., and Bertoncini, J. (**1997**). "Organization and discrimination of repeating sound sequences by newborn infants," J. Acoust. Soc. Am. **102**, 2945–2953.
McKinley, R. C., and Weber, D. L. (**1994**). "Detection and recognition of repeated tones and tonal patterns," J. Acoust. Soc. Am. **95**, 2642–2651.

Neff, D. L. (**1995**). "Signal properties that reduce masking by simultaneous, random-frequency maskers," J. Acoust. Soc. Am. **98**, 1909–1920.
Neff, D. L., and Dethlefs, T. M. (**1995**). "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," J. Acoust. Soc. Am. **98**, 125–134.
Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.
Oh, E. L., and Lutfi, R. A. (**1998**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **104**, 3489–3499.
Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). "Children's detection of pure-tone signals with random multitone maskers," J. Acoust. Soc. Am. **109**, 2888–2895.
Rose, M. M., and Moore, B. C. (**1997**). "Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **102**, 1768–1778.
Sussman, E., Wong, R., Horvath, J., Winkler, I., and Wang, W. (**2007**). "The development of the perceptual organization of sound by frequency separation in 5-11-year-old children," Hear. Res. **225**, 117–127.
Swets, J. A., Shipley, E. F., McKey, M. F., and Green, D. M. (**1959**). "Multiple observations of signals in noise," J. Acoust. Soc. Am. **31**, 514–521.
Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
Watson, C. S., Wroton, H. W., Kelly, W. J., and Benbassat, C. A. (**1975**). "Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals," J. Acoust. Soc. Am. **57**, 1175–1185.
Werner, L. A. (**1996**). "The development of auditory behavior (or what the anatomist and physiologists have to explain," Ear Hear. **17**, 438–446.
Wightman, F., Kistler, D., and Brungart, D. (**2006**). "Informational masking of speech in children: Auditory-visual integration," J. Acoust. Soc. Am. **119**, 3940–3949.
Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). "Children's detection of pure-tone signals: Informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305.
Wightman, F. L., and Kistler, D. J. (**2005**). "Informational masking of speech in children: Effects of ipsilateral and contralateral distracters," J. Acoust. Soc. Am. **118**, 3164–3176.
Winkler, I., Kushnerenko, E., Horvath, J., Ceponiene, R., Fellman, V., Huotilainen, M., Naatanen, R., and Sussman, E. (**2003**). "Newborn infants can organize the auditory world," Proc. Natl. Acad. Sci. U.S.A. **100**, 11812–11815.
Yost, W. A. (**1991**). "Auditory image perception and analysis: The basis for hearing," Hear. Res. **56**, 8–18.
Yost, W. A. (**1997**). in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. A. Gilkey and T. R. Anderson (Lawrence Erlbaum Associates Inc., Mahwah, NJ), pp. 329–347.

Maximal azimuthal resolution needed in measurements of head-related transfer functions

Xiao-Li Zhong and Bo-Sun Xie[a)]
Physics Department, Acoustic Laboratory, South China University of Technology, Guangzhou 510641, China

(Received 23 March 2008; revised 2 February 2009; accepted 3 February 2009)

The head-related transfer function (HRTF) is a continuous function of sound source position. Measurement of the HRTF can only be undertaken at discrete positions in space, however. Determination of angular resolutions so as to reconstruct HRTFs at unmeasured positions has been an open problem. Azimuthal Fourier analysis was proposed to analyze the variation in the HRTF in each elevation plane. As a result, an azimuthal sampling theorem and a corresponding interpolation formula were derived. It was proved that the maximal azimuthal resolution of measurements is $360°/(2Q+1)$, where Q represents the highest order in the truncation of the azimuthal Fourier expansion of the HRTF. The maximal azimuthal resolutions for the HRTF with and without arrival time correction were investigated. Results show that the arrival time correction can reduce the burden of measurements, since a larger azimuthal resolution is possible without introducing obvious interpolation error. A psychoacoustic experiment was also conducted to evaluate the proposed theorem. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3087433]

PACS number(s): 43.66.Pn, 43.60.Ac [JCM] Pages: 2209–2220

I. INTRODUCTION

Head-related transfer functions (HRTFs) reflect the overall filtering effects caused by anatomical structures such as the head, torso, and pinna. Because HRTFs contain all primary sound localization cues (Blauert, 1997; Kulkarni and Colburn, 1998; Wenzel *et al.*, 1993), the nature and use of HRTFs are important for understanding binaural hearing and in the implementation of virtual auditory displays (Shinn-Cunningham, 1998; Xie, 2008).

Although calculation is a way to obtain HRTFs (Algazi *et al.* 2002; Katz, 2001), measurement remains a common approach. To date, measurements have been undertaken on artificial heads and human subjects by various research groups (Algazi *et al.*, 2001; Blauert *et al.*, 1998; Bovbjerg *et al.*, 2000; Gardner and Martin, 1995; Genuit and Xiang, 1995; Middlebrooks *et al.*, 1989; Møller *et al.*, 1995; Riederer, 1998; Wightman and Kistler, 1989; Xie *et al.*, 2007).

Naturally, HRTFs vary as a continuous function of sound source position. Only the HRTFs at discrete positions can be measured in practice, representing a sampling of the continuous three-dimensional space for sound source positions. Differing spatial resolutions have been used by different research groups. For example, in the horizontal plane, the azimuthal resolutions were in a range 0.9°–22.5° in the references above. In order to obtain HRTFs at unmeasured positions, a number of interpolation methods have been proposed (Biscainho *et al.*, 2002; Nishino *et al.*, 1999; Wenzel and Foster, 1993; Wightman *et al.*, 1992). In view of achieving good interpolation performance, measurement with a high spatial resolution is preferable. However, because dense measurement is time consuming and laborious, it is inconvenient to measure HRTFs with a very fine spatial resolution. Then, the problem of how to determine an appropriate spatial resolution in HRTF measurement arises.

Minnaar *et al.* (2005) analyzed the spatial resolutions needed in HRTF measurement by examining the audible error introduced by the linear interpolation of two neighboring measurements. They found that different spatial regions required different spatial resolutions. For example, in the horizontal plane, an azimuthal resolution of 16° was sufficient except for the lateral region where an azimuthal resolution of 8° was required. Since an audible criterion was used, the resulting spatial resolutions could not ensure the interpolation accuracy in mathematics.

Inspired by the reciprocal principle of acoustics, Ajdler *et al.* (2005) deduced an azimuthal sampling theorem in the horizontal plane from the theory of sound-field pickup, in which the maximal azimuthal resolution was directly related to the bandwidth of the HRTF. This theorem is only an approximation, however, since it was not derived from measured HRTFs and did not fully account for the complex sound wave diffraction introduced by the anatomical system.

According to the signal processing theory, the maximal azimuthal resolution needed to reconstruct a continuous signal is determined by the complexity of the signal itself. Analyzing the spatial variation of the HRTF is therefore a way to determine the spatial resolution of the HRTF. Preliminary work relating the spatial variation of the HRTF to its reconstruction was done by Evans *et al.* (1998), but no explicit discussion of the spatial resolution of the HRTF was given. Moreover, since Evans *et al.* (1998) expanded HRTFs by surface spherical harmonics, there are some limitations on the spatial resolution of the HRTF in order to estimate the expansion coefficients effectively using the Gauss–Legendre quadrature formula.

[a)]Author to whom correspondence should be addressed. Electronic mail: phbsxie@scut.edu.cn

In the current study, the azimuthal variation of the HRTF is investigated using azimuthal Fourier analysis. As a result, an azimuthal sampling theorem of HRTF is obtained, which is similar to the famous Nyquist sampling theorem, and a corresponding interpolation formula is derived. Unlike the spherical-harmonic expansion, which deals with elevation and azimuth simultaneously, the azimuthal Fourier analysis of the HRTF focuses only on HRTF variation with azimuth in a given elevation plane. The resulting azimuthal sampling theorem allows different azimuthal resolutions in different elevation planes. The basic idea of the azimuthal Fourier analysis appeared in Zhong and Xie (2005), which analyzed the spatial variation of the HRTF from KEMAR. The present paper sets out a detailed exposition of the basic theory of the azimuthal Fourier analysis. Analyses of the HRTFs from KEMAR and humans under the original and the arrival time correction representation are elaborated. Finally, recommendations are given on the maximal azimuthal resolution for different elevations based on the statistical results of 52 human subjects.

II. FUNDAMENTALS

The center of the head is chosen as the origin of coordinates. The distance from a sound source to the origin is denoted by r. The sound source direction is specified by the azimuth θ from 0° to 360° and the elevation ϕ from –90° to 90°, where ϕ=0° and 90°, respectively, correspond to the horizontal and vertical plane. In the horizontal plane, θ=0° is the front and θ=90° is the right.

In the free field, HRTF is defined as

$$H(r,\theta,\phi,f)=\frac{P(r,\theta,\phi,f)}{P_0(r,f)}, \tag{1}$$

where P denotes the sound pressure at the left or the right ear caused by a sound source, and can be measured anywhere from the eardrum to the entrance of the ear canal; P_0 denotes the sound pressure at the origin with the head absent, and f is the frequency of sound waves. Generally, HRTF is a complicated function of the sound source position (including sound source direction and distance) and the frequency. In the far field, however, when the distance from the head center to the sound source exceeds 1.0–1.2 m, the HRTF is asymptotically independent of distance. Also, the HRTF is closely related to anatomical features which differ from person to person, so that the HRTF is individual-dependent. The counterpart of the HRTF in the time domain is called the head-related impulse response (HRIR), which is the inverse Fourier transform of the HRTF.

A. Azimuthal Fourier analysis and azimuthal sampling theorem

In the far field, the HRTF is a continuous function of azimuth with a period of 2π for a given elevation. It can therefore be expanded as an azimuthal Fourier series as

$$H(\theta,f)=\sum_{q=-\infty}^{+\infty} d_q(f)e^{iq\theta}, \quad d_q(f)=\frac{1}{2\pi}\int_0^{2\pi} H(\theta,f)e^{-iq\theta}d\theta, \tag{2}$$

where $H(\theta,f)$ denotes the azimuthal HRTF at a given elevation, $d_q(f)$ is the weight coefficient, and q the expansion order. In Eq. (2), the azimuthal HRTF is decomposed as a linear combination of infinite weighted azimuthal harmonics. If the azimuthal harmonic expansion in Eq. (2) is convergent, Eq. (2) can be truncated as

$$H(\theta,f)=\sum_{q=-Q}^{+Q} d_q(f)e^{iq\theta}, \tag{3}$$

where Q is the truncation order, above which the higher-order harmonics are neglected because their contribution is insignificant. The azimuthal HRTF then consists of $(2Q+1)$ weighted azimuthal harmonics, and is completely determined by their $(2Q+1)$ weight coefficients. By sampling the azimuthal HRTF from azimuth 0 to 2π at M positions with an azimuthal resolution $2\pi/M$, we have

$$H(\theta_m,f)=\sum_{q=-Q}^{+Q} d_q(f)e^{iq\theta_m}, \quad \theta_m=\frac{2\pi m}{M}, \quad m=0,1,\ldots,(M-1), \tag{4}$$

where θ_m is the measured azimuth, and $H(\theta_m,f)$ is the measured HRTF at azimuth θ_m. If the number of azimuthal measurements $M\geq(2Q+1)$, then $(2Q+1)$ weight coefficients can be obtained by resolving the M linear equations in Eq. (4) using the orthogonality and completeness properties of $\{\exp(iq\theta_m)\}$:

$$d_q(f)=\frac{1}{M}\sum_{m=0}^{M-1} H(\theta_m,f)e^{-iq\theta_m}, \quad |q|\leq Q,$$

$$d_q(f)=0, \tag{5}$$

$$Q<|q|\leq(M-1)/2 \quad \text{if } M \text{ is odd},$$

$$-(M/2)\leq q<-Q \quad \text{and} \quad Q<q\leq(M/2)-1 \quad \text{if } M \text{ is even}.$$

Substitution of Eq. (5) into Eq. (3) yields

$$H(\theta,f)=\frac{1}{M}\sum_{m=0}^{M-1} H(\theta_m,f)\frac{\sin\left[\left(Q+\frac{1}{2}\right)(\theta-\theta_m)\right]}{\sin\left(\frac{\theta-\theta_m}{2}\right)}. \tag{6}$$

Equation (6) allows the HRTFs at unmeasured azimuths to be reconstructed from the M measured HRTFs. Note that the derivation of Eqs. (5) and (6) supposes that the number of azimuthal measurements $M\geq(2Q+1)$. Therefore, $(2Q+1)$ is defined as the minimal number of azimuthal measurements (MNAM) for the HRTF, and correspondingly the maximal azimuthal resolution is 360°/MNAM.

In conclusion, for a given elevation, the azimuthal HRTF can be decomposed as a linear combination of

weighted azimuthal harmonics. When the number of azimuthal measurements is not less than the MNAM, or equivalently when the azimuthal resolution is not larger than the maximal azimuthal resolution 360°/MNAM, the HRTFs at unmeasured azimuths can be reconstructed from measured HRTFs at discrete azimuths using Eq. (6). This is the azimuthal sampling theorem of the HRTF, which is similar to the famous Nyquist theorem with respect to time.

Because the azimuthal variation of the HRTF is unlikely to be the same for different frequencies, the truncation order Q, and hence the MNAM, depends on frequency.

B. Determination of MNAM

In order to determine Q and then the MNAM, the relative power of the sth-order azimuthal harmonic term against the total power is defined as

$$\eta_s(f) = \frac{|d_s(f)|^2 + |d_{-s}(f)|^2}{\sum_{q=-\infty}^{+\infty} |d_q(f)|^2}, \quad s = 1,2,3,\ldots. \tag{7}$$

In Eq. (7), the denominator is an infinite sum over weight coefficients which should be calculated from the azimuthal HRTF, as shown in Eq. (2). Unfortunately, the azimuthal HRTF within the audible frequency range is not available, since only HRTFs at discrete positions can be obtained by measurements. However, if a HRTF database exists which contains sufficient measured HRTFs and satisfies the azimuthal sampling theorem of the HRTF, then the coefficients can be calculated by Eq. (5). Equation (7) can then be rewritten as

$$\eta_s(f) = \frac{|d_s(f)|^2 + |d_{-s}(f)|^2}{\sum_{q=-Q_1}^{+Q_2} |d_q(f)|^2}, \tag{8}$$

where (Q_1+Q_2+1) is the number of azimuthal measurements M. To evaluate the contribution of the preceding S-order azimuthal harmonic terms, the relative power ratio η is defined as

$$\eta = \frac{\sum_{q=-S}^{+S} |d_q(f)|^2}{\sum_{q=-Q_1}^{+Q_2} |d_q(f)|^2}, \quad \begin{array}{l} S = 1,2,\ldots,\min(Q_1,Q_2). \\ Q_1 = Q_2 = (M-1)/2 \ \text{ if } M \text{ is odd.} \\ Q_1 = M/2, Q_2 = (M/2)-1 \ \text{ if } M \text{ is even.} \end{array} \tag{9}$$

The steps for determining Q and hence the MNAM for each elevation are as follows.

(1) At each discrete frequency point whose value is determined by the frequency sampling rate, calculate η for $S=1,2,\ldots,\min\ (Q_1,\ Q_2)$ until $S=S_1$ when the η reaches a power criterion (such as 0.95) that is chosen according to the accuracy desired. Then, S_1 is regarded as the truncation order Q above which the contribution of higher-order terms is negligible. According to the azimuthal sampling theorem of the HRTF, $(2Q+1)=(2S_1+1)$ is the MNAM.

(2) If $S_1<\min(Q_1,Q_2)$, then $(2S_1+1)<2\ \min(Q_1,Q_2)+1 \leq (Q_1+Q_2+1)$, where the first term is the MNAM and the last term is the number of azimuthal measurements. In this case, because the MNAM is less than the number of azimuthal measurements, the number of azimuthal measurements is sufficient with some redundancy under the chosen power criterion. Otherwise, if $S_1=\min(Q_1,Q_2)$, which means the truncation rests on the harmonic term with the highest order that the HRTFs employed can provide, then it is difficult to judge whether the number of azimuthal measurements just suffices or not. In this case, the number of azimuthal measurements analyzed is regarded as insufficient.

(3) The MNAM calculated by the steps above is just for each discrete frequency point. For a given frequency range, such as the audible frequency range, the maximum among the MNAMs of the discrete frequency points included in the frequency range is chosen as the MNAM of the frequency range. In this paper, MNAM of the audible frequency range is denoted by $\text{MNAM}_{\text{audio}}$.

C. Arrival time correction representation of the HRTF

Section II A deals with the original HRTF, and Eq. (6) is the azimuthal interpolation equation for the original HRTF. Since the HRIR is the time domain version of the HRTF, Eq. (6) can also be regarded as an azimuthal interpolation equation for the original HRIR.

Each original HRIR contains a position-dependent onset delay, relating roughly to the arrival time of sound waves from the sound source to the ear. Some studies found that if the arrival time was corrected prior to interpolation, i.e., the arrival time for each position was made synchronous by shifting the onset time of the HRIR, then the interpolation performance of the HRIR or HRTF could be improved (Evans *et al.*, 1998; Matsumoto *et al.*, 2004). This motivates correction for the arrival time in the original HRTF in the azimuthal Fourier analysis.

There are several ways to determine the time displacement. In the work of Evans *et al.* (1998), the time displacement was chosen as a time segment from the beginning of the HRIR to four samples before the instant when the HRIR first exceeded an onset threshold of 0.015. An alternative method is used here. For a given elevation plane, the time displacement is chosen as the time value that maximizes the cross-correlation function between the HRIR at azimuth 0° and the HRIR at each azimuth θ_m. This is equivalent in the frequency domain to maximizing the cross-spectra function. Moreover, modifying the arrival time in the HRIR is equivalent to transforming the original HRTF to its time-corrected version H_τ as

$$H_\tau(\theta,f) = H(\theta,f)\exp[-j2\pi f\tau(\theta)], \tag{10}$$

where $\tau\ (\theta)$ is the time displacement. A positive time displacement means incorporating a time delay to the original

TABLE I. The number of azimuthal measurements for KEMAR and calculated $MNAM_{audio}$ to reconstruct HRTFs at unmeasured positions up to 20 kHz.

			Calculated MNAM				
			$HRTF_{original}$		$HRTF_{correction}$		
					H_τ		
Elevation	The number of azimuthal measurements for KEMAR	Azimuthal resolution[a]	0.99	0.95	0.99	0.95	Time displacement
−30°	60	6°	57	53	49	21	37
−20°	72	5°	63	57	35	17	19
−10°	72	5°	63	57	31	17	15
0°	72	5°	65	59	43	21	13
10°	72	5°	67	59	59	25	15
20°	72	5°	65	59	57	23	17
30°	60	6°	57	53	39	19	13
50°	45	8°	45	41	31	21	27
60°	36	10°	31	29	19	17	5
70°	24	15°	23	23	17	15	5
80°	12	30°	11	11	11	11	3

[a]Gardner and Martin (1995).

HRTF, whereas a negative value means incorporating a time advance.

Below are the steps for obtaining the interpolated HRTFs under the arrival time correction representation.

(1) At a given elevation, suppose there are M original HRTFs measured with an azimuthal resolution of $(360°/M)$, among which the first sampled azimuth is 0°. Then, M time displacements can be estimated by calculating the delay that maximizes the cross-spectra function between the HRTF at azimuth 0° and the HRTF at azimuth θ_m. Note the time displacement is zero for azimuth 0°. Next, as shown in Eq. (10), shifting of the original HRTFs according to the calculated time displacements yields the $H_\tau s$, which have the same onset time as in the original HRTF at azimuth 0°.
(2) As with the original HRTF, the azimuthal H_τ at a given elevation is also a periodic function of azimuth with a period of 2π, so that the methods proposed in Secs. II A and II B can also be applied. The MNAM for the azimuthal H_τ can be calculated and the H_τ at unmeasured azimuths can be obtained using Eq. (6).
(3) The methods in Secs. II A and II B can also be applied to the azimuthal function of time displacement τ in a given elevation for its periodicity. The MNAM for the time displacement can be calculated and the time displacements at unmeasured azimuths can be interpolated by

$$\tau(\theta) = \frac{1}{M}\sum_{m=0}^{M-1} \tau(\theta_m)\frac{\sin\left[\left(Q' + \frac{1}{2}\right)(\theta - \theta_m)\right]}{\sin\left(\frac{\theta - \theta_m}{2}\right)}, \tag{11}$$

where $\tau(\theta_m)$ is the measured time displacement obtained by step (1), and Q' is the truncation order of the azimuthal function of time displacement, which may be different from the truncation order of the azimuthal H_τ.
(4) For the same azimuth, add the interpolated time displacement in step (3) back to the corresponding interpolated H_τ in step (2) to obtain the overall HRTF under the arrival time correction representation.

As with the original HRTF, a power criterion is used to determine the truncation order of the azimuthal harmonics of the azimuthal H_τ in step (2). For the truncation of the azimuthal function of time displacement, a time criterion such that truncation error is less than 5 μs is used to ensure that the largest error of the interaural time difference does not exceed 10 μs.

For the arrival time correction representation of HRTF, the MNAM needed to reconstruct the overall HRTFs at unmeasured azimuths is determined by the larger one of the MNAM for the H_τ and the MNAM for the time displacement. It will be shown in Sec. III C that the final MNAM is always determined by the former, that is, the MNAM for the H_τ. In the following, the HRTFs under the original and the arrival time correction representation are denoted by $HRTF_{original}$ and $HRTF_{correction}$, respectively.

III. RESULTS FOR MNAM

A. HRTF database

Two sets of HRTF database are used.

(1) The HRTF database for KEMAR from the MIT media Laboratory (Gardner and Martin, 1995).

Since KEMAR is designed according to the mean anatomical size of the Western population, results from KEMAR HRTF are representative in a sense of the mean (Burkhard and Sachs, 1975). Moreover, the measurement error of KEMAR is small as KEMAR is not liable to movement during measurements. The far-field HRTF data from KEMAR with small pinna (DB-061) are therefore used, as shown in Table I. For each measured position, the database contains a 512-

TABLE II. The number of azimuthal measurements for the HRTF database containing 52 human subjects and calculated mean $MNAM_{audio}$ to reconstruct HRTFs up to 20 kHz. Corresponding standard deviations are in parentheses.

Elevation	The number of azimuthal measurements	Azimuthal resolution	Calculated mean $MNAM_{audio}$ and standard deviation with power criterion 0.95	
			For $HRTF_{original}$	For $HRTF_{correction}$
−30°	72	5°	60 (±4.8)	37 (±11.3)
−15°	72	5°	63 (±3.9)	29 (±10.0)
0°	72	5°	64 (±3.9)	36 (±10.3)
15°	72	5°	62 (±4.8)	29 (±5.7)
30°	72	5°	58 (±4.9)	28 (±7.7)
45°	72	5°	52 (±7.4)	24 (±8.0)
60°	36	10°	34 (±1.3)	16 (±3.2)
75°	24	15°	23 (±1.4)	14 (±3.6)

tap HRIR with a frequency sampling rate of 44.1 kHz, from which a 512-tap HRTF can be obtained by the discrete-time Fourier transform.

(2) HRTF database of human subjects(Xie *et al.*, 2007).

To obtain statistical results of human subjects, a HRTF database established recently by our laboratory was used, which contains HRTFs of 52 subjects, 26 male and 26 female. For each subject, the far-field HRTFs ($r=1.5$ m) were measured at 493 spatial positions, including eight elevation planes. Azimuthal resolutions are shown for each elevation plane in column 3 of Table II. For each measured position, a pair of 512-tap HRIRs for the left and right ears with a frequency sampling rate of 44.1 kHz was obtained. The method in Sec. II can be applied to either left-ear or right-ear HRTFs. For simplicity, only the data from the left ear are used in the following analysis.

B. MNAM for the $HRTF_{original}$

Figure 1(a) shows how the relative power contribution of each azimuthal harmonic of the original HRTF varies with frequency in the horizontal plane. Figure 1(b) is for the H_τ, which will be discussed in Part C. In Fig. 1(a), the relative contribution of lower-order azimuthal harmonics, which correspond to the smooth azimuthal variation of the HRTF, is significant at low frequencies when the wavelength is large. As a result, the truncation order is low and the MNAM is small. With increasing frequency, the interference of sound waves caused by the head and pinna becomes significant. As a result, the HRTF varies with azimuth rapidly at high frequencies, and the relative contribution of higher-order azimuthal harmonics increases. For example, the relative contributions of the 20th-order azimuthal harmonic are 3.2×10^{-7} at 172 Hz and 1.2×10^{-1} at 15.4 kHz. Therefore, the truncation order is likely to be high, so that the MNAM is large at high frequencies. Moreover, Fig. 1(a) shows that, at each frequency sampling point, the relative contribution of the azimuthal harmonics demonstrates a clear downward trend with increasing expansion order, verifying the hypothesis in Sec. II A that the azimuthal harmonic expansion of the HRTF is convergent.

The results for other elevation planes are similar, but the relative contribution of high-order azimuthal harmonics declines gradually as the elevation deviates from the horizontal plane. For example, the relative contributions of the 20th-order azimuthal harmonic at 15.4 kHz are 1.2×10^{-1}, 5.1×10^{-2}, and 2.3×10^{-3} for elevation 0°, 20°, and 50° in turn.

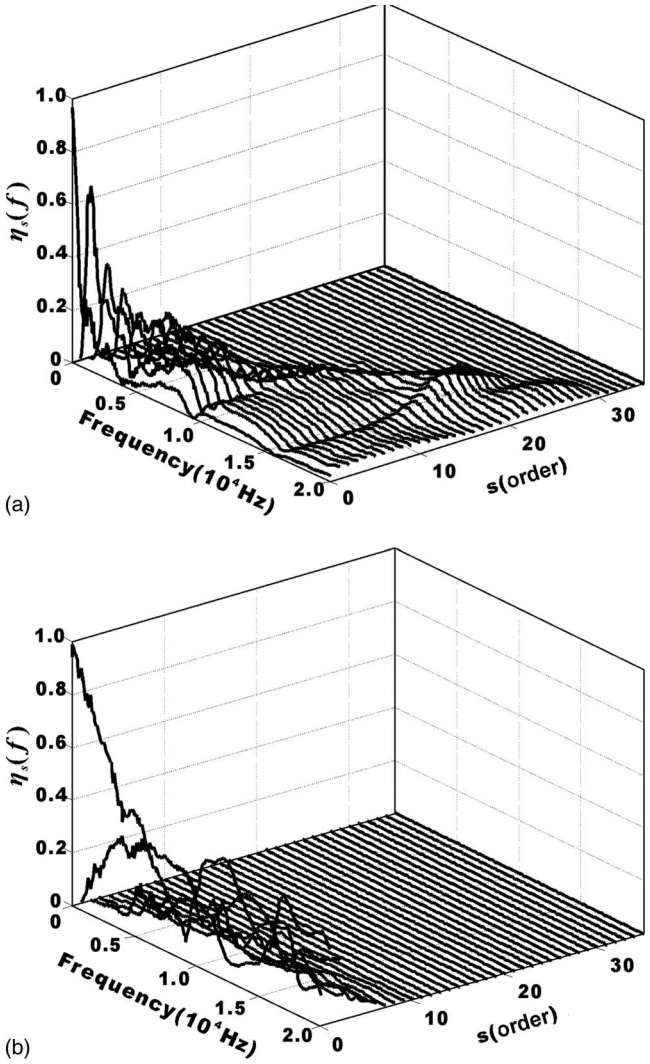


FIG. 1. In the horizontal plane, the relative contribution of each azimuthal harmonic of KEMAR is shown, with (a) for the $HRTF_{original}$ and (b) for the H_τ calculated from Eq. (8).

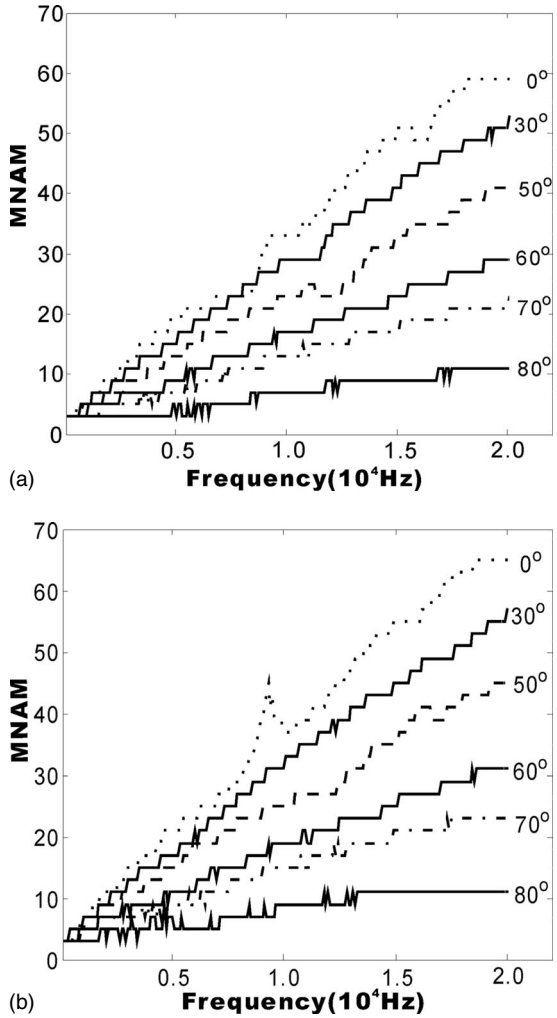


FIG. 2. Calculated MNAM of KEMAR for different elevation planes. The abscissa is frequency, and the ordinate is the MNAM of the power criterion (a) 0.95 and (b) 0.99. The MNAMs for elevations 10° and 20° (omitted) are very similar to that of elevation 0°.

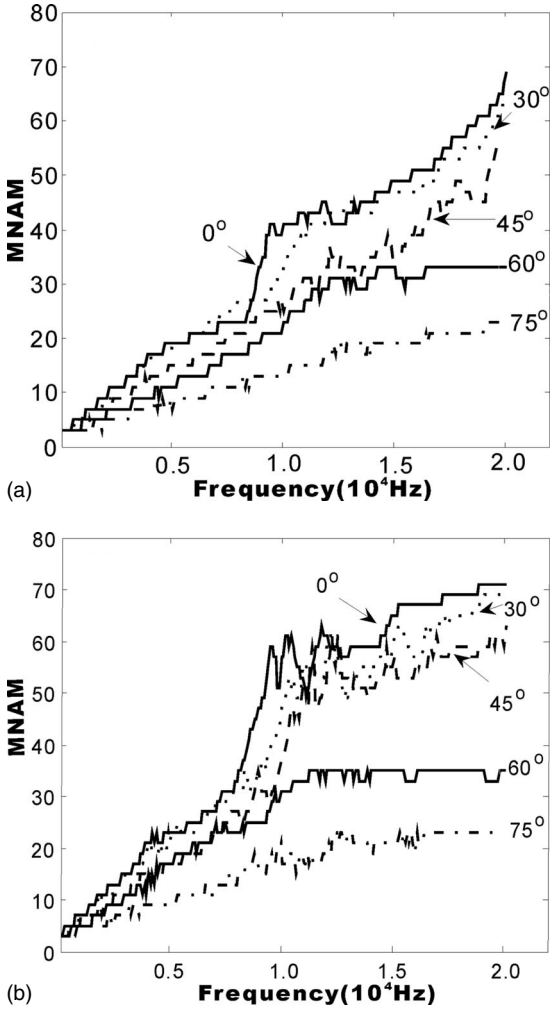


FIG. 3. Calculated MNAM for the human subject 23 in different elevation planes with the power criterion (a) 0.95 and (b) 0.99. The MNAM for elevation 15° is very similar to that of elevation 30°, and is omitted for clarity.

This is because the variation in the HRTF with azimuth smoothens gradually with elevation shifting from the horizontal plane.

Figure 2 shows the MNAM for KEMAR as a function of frequency and elevation. The features in this plot are as follows.

(1) Given an elevation plane and a frequency sampling point, the MNAM with criterion 0.99 in Fig. 2(b) is always larger than that with criterion 0.95 in Fig. 2(a), because the former is stricter than the latter in accuracy.
(2) Given an elevation plane, the MNAM of each frequency point increases stepwise with increasing frequency, although a few exceptions may exist. Therefore, the MNAM of a frequency range, which is the maximum among the MNAMs for all discrete frequency points in the range, is likely to be the MNAM of the high-limit frequency sampling point in that range.
(3) Given a frequency sampling point, the MNAM decreases gradually as elevation deviates from the horizontal plane. This is because the azimuthal variation of the HRTF smoothes gradually with elevation shifting from the horizontal plane.

The HRTF of the audible frequency range (up to 20 kHz) has been widely measured and employed. Therefore, the $\mathrm{MNAM}_{\mathrm{audio}}$ needed to reconstruct the HRTFs for KEMAR at unmeasured azimuths up to 20 kHz were calculated according to step (3) in Sec. II B. Results are listed in columns 4 and 5 of Table I. Comparison with the numbers of azimuthal measurements in column 2 of Table I shows that, for the criterion 0.99, the numbers of azimuthal measurements in the MIT database are sufficient except for elevations 50°, 70°, and 80°, while for the power criterion 0.95 only the numbers of azimuthal measurements for elevations 70° and 80° are likely to be insufficient.

The methods in Secs. II A and II B can also be implemented on human HRTFs. Figure 3 shows how the MNAM calculated from the left-ear HRTFs of a typical subject (No. 23 in the database) varies with elevation and frequency. Compared with Fig. 2, a similar tendency is observed, i.e., the MNAM in a given elevation increases with increasing frequency, and for a given frequency sampling point the MNAM decreases gradually with elevation deviating from the horizontal plane.

Calculations also indicate that the MNAM for humans is

TABLE III. Suggested azimuthal measurement number for human HRTF measurement with azimuthal resolutions in parentheses, which are determined with arrival time correction.

Elevation	Calculated mean $MNAM_{audio}$ and standard deviation with criterion 0.95	Suggested azimuthal measurement number needed to reconstruct HRTF up to 20 kHz with a confidence level of 95%	Suggested azimuthal measurement number needed to reconstruct HRTF up to 20 kHz according to conventional selection
−30°	37 (±11.3)	60 (6°)	72 (5°)
−15°	29 (±10.0)	60 (6°)	72 (5°)
0°	36 (±10.3)	60 (6°)	72 (5°)
15°	29 (±5.7)	40 (9°)	72 (5°)
30°	28 (±7.7)	45 (8°)	72 (5°)
45°	24 (±8.0)	40 (9°)	72 (5°)
60°	16 (±3.2)	24 (15°)	24 (15°)
75°	14 (±3.6)	24 (15°)	24 (15°)

always larger than that for KEMAR under the same conditions, including elevation, frequency point, and power criterion. This is because the azimuthal variation of human HRTFs is more complex than that of KEMAR, since humans have finer anatomical appearance and are liable to slight movements during measurements. In view of the relatively large error in human HRTF measurements, the criterion 0.95 is used in the following analysis of human HRTFs.

According to step (3) in Sec. II B, the mean $MNAM_{audio}$ and the standard deviations σ for 52 human subjects were calculated with criterion 0.95. The results are listed in column 4 of Table II, which shows that almost all the numbers of azimuthal measurements in column 2 are larger than the upper-limit values of the calculated mean $MNAM_{audio}$ at 95% confidence level (that is, the values of (mean+2σ)) except for a slight insufficiency at elevations 60° and 75°. Consequently, the numbers of azimuthal measurements used in our HRTF database are sufficient to reconstruct the HRTFs at unmeasured azimuths up to 20 kHz in almost all measured elevation planes.

C. MNAM for the $HRTF_{correction}$

In the horizontal plane, KEMAR HRTFs at 72 measured azimuths can be decomposed into 72H_τ and 72 time displacements according to step (1) in Sec. II C. As discussed in step (2) in Sec. II C, the spatial Fourier analysis in Sec. II A can also be applied to the azimuthal H_τ. The relative power contribution of each azimuthal harmonic of the azimuthal H_τ evaluated from Eq. (8) is shown in Fig. 1(b). Upon comparing Fig. 1(b) with Fig. 1(a), it is apparent that the arrival time correction smooths the spatial variation of the HRTF by decreasing the coefficients of high-order azimuthal harmonics. This explains why the arrival time correction improves the interpolation performance reported by Evans *et al.* (1998) and Matsumoto *et al.* (2004).

According to steps (2) and (3) in Sec. II C, the $MNAM_{audio}$ for KEMAR H_τ with criterions 0.99 and 0.95, and for corresponding time displacement, are listed in columns 6–8 of Table I, respectively. The data show that the $MNAM_{audio}$ for the H_τ is always larger than that for the time displacement in each elevation plane, except for elevations −30°, −20°, and 50° with criterion 0.95. Therefore, the $MNAM_{audio}$ for the $HRTF_{correction}$ is almost determined by the $MNAM_{audio}$ for the H_τ, as mentioned in Sec. II C. In Table I, comparison of the $MNAM_{audio}$ values for the $HRTF_{original}$ listed in columns 4 and 5 with those for the $HRTF_{correction}$ listed in columns 6 and 7 reveals that the latter are smaller. For example, in the horizontal plane, $MNAM_{audio}$ decreases from 65 for the $HRTF_{original}$ to 43 for the $HRTF_{correction}$ with criterion 0.99, and 59 for the $HRTF_{original}$ to 21 for the $HRTF_{correction}$ with criterion 0.95.

The arrival time correction is also applied to the HRTFs of 52 human subjects, and the mean $MNAM_{audio}$ (with standard deviation) is listed in column 5 of Table II. The mean $MNAM_{audio}$ for human $HRTF_{correction}$ is slightly larger than that for KEMAR due to the finer anatomical detail and larger measurement error of humans. As with KEMAR, the arrival time correction reduces the mean $MNAM_{audio}$ nearly by half. This suggests that the sampling positions needed in human HRTF measurement could be halved by using the arrival time correction.

D. Discussion

Calculations show that, for the same HRTF data of humans, the mean $MNAM_{audio}$ for the $HRTF_{correction}$ is about half of that for the $HRTF_{original}$. For the both representations, the overall HRTF including magnitude and phase can be reconstructed exactly, if the azimuthal sampling theorem of the HRTF is satisfied. The MNAM used in HRTF measurement should therefore be determined according to the results with the arrival time correction in view of experimental efficiency. For humans, the mean $MNAM_{audio}$ (up to 20 kHz) with the arrival time correction is listed in column 5 of Table II and rewritten in column 2 of Table III. To make practical measurement convenient and satisfy the sampling theorem, the suggested number of azimuthal measurements listed in column 3 of Table III is chosen as an integer that divides 360° exactly and is equal to or a little larger than the upper-limit value of the mean $MNAM_{audio}$ at 95% confidence level; the corresponding azimuthal resolution is also given. Since con-

TABLE IV. Conditions for the interpolation in the horizontal plane.

HRTF set	KEMAR		Subject 23[(1)]	
HRTF representation	Original	Arrival time correction	Original	Arrival time correction
Calculated $MNAM_{audio}$ with criterion 0.95	59	21	69	31
Division pattern	1	2	1	1
Reconstructed frequency range (kHz)	≤11	≤20	≤9	≤20
SDR_{mean} (dB)	21.4	17.3	19.6	16.7

[a]Results for subject 23 are used to demonstrate the typical interpolation performance of humans. The reconstructed frequency range and the SDR_{mean} may vary slightly with subject.

ventional HRTF measurement is usually performed at intervals of multiples of 5°, an alternative suggestion is given in column 4 of Table III.

Evans *et al.* (1998) expanded measured HRTFs using surface spherical harmonics as basis functions. Because the surface spherical harmonics vary as a function of both azimuth and elevation, some limitations are imposed on sampling positions in HRTF measurements. In the work of Evans *et al.* (1998), to obtain the spherical-harmonic coefficients up to and including l-order from measured HRTFs at discrete positions exactly and with mathematical efficiency, measurements should be carried out at L $(L \geq l+1)$ specific elevations determined by the Gauss–Legendre quadrature, including some low elevations such as $-82.5°$ for which the measurement is difficult. Moreover, measurements at $2L$ equally-spaced azimuths are needed in each elevation plane measurement, which is difficult and inefficient, particularly at high elevation planes. Most HRTF databases available at present are unsuitable for the surface spherical-harmonic expansion. In contrast, for the azimuthal Fourier expansion of the HRTF, only uniform azimuthal measurements are needed at each elevation. However, only the HRTFs at unmeasured azimuths can be reconstructed by the azimuthal Fourier analysis, and other methods should be developed to reconstruct the HRTFs at unmeasured elevations.

Using the theory of sound-field pickup, Ajdler *et al.* (2005) deduced that the maximal azimuthal resolution for HRTF measurement in the horizontal plane was about 5° for the audible frequency range, or equivalently that $MNAM_{audio}$ was 72. In Tables I and II, the $MNAM_{audio}$ for KEMAR in the horizontal plane are 65 and 59 with criterions 0.99 and 0.95, respectively, and for human subjects it is 64 with criterion 0.95. It is encouraging that our result is consistent with that of Ajdler *et al.* (2005). However, the method proposed by Ajdler *et al.* (2005) is unable to solve the situation with the arrival time correction. Moreover, only the azimuthal sampling in the horizontal plane was discussed by Ajdler *et al.* (2005).

IV. INTERPOLATION PERFORMANCE AND ANALYSIS

A. Procedure

Validity of the azimuthal sampling theorem and the resulting interpolation formula Eq. (6) are evaluated using the HRTFs in the horizontal plane, where HRTF varies with azimuth most rapidly. As shown in Tables I and II, HRTFs were measured at 72 azimuthal positions with an azimuthal resolution of 5° in the horizontal plane. Those positions are divided into two subsets: a calculation subset serving as input data to the interpolation and a comparison subset that will be predicted based on interpolation. Thus, for each position in the comparison subset, HRTF obtained via interpolation will be compared to the measured HRTF.

Two kinds of division patterns are used.

(1) Pattern 1: The 36 positions from azimuth 0°–350° at an interval of 10° make up the calculation subset, and the remainder 36 positions make up the comparison subset.
(2) Pattern 2: The 24 positions from azimuth 0°–345° at an interval of 15° make up the calculation subset, and the remainder 48 positions make up the comparison subset.

The two kinds of division methods are used to simulate two kinds of the measurement numbers: 36 in pattern 1 and 24 in pattern 2.

According to the azimuthal sampling theorem, the HRTF of a frequency range can be reconstructed by interpolation when the number of azimuthal measurements is not less than the MNAM of that range. In practice, the number of azimuthal measurements is always selected as equal to or just a little larger than the MNAM of the frequency range, which is also the basic principle in selecting division patterns. The two possible cases which may arise in selecting division patterns are as follows.

(1) If the $MNAM_{audio}$ in Table IV is not larger than 36, the division pattern in which the number of azimuthal measurements (the number of positions in the calculation subset) is equal to or a little larger than the $MNAM_{audio}$ is selected. For example, in the horizontal plane, the $MNAM_{audio}$ for KEMAR $HRTF_{correction}$ is 21, and pattern 2 is used with the azimuthal measurement number of 24, whereas for subject 23, the $MNAM_{audio}$ for the $HRTF_{correction}$ is 31, and pattern 1 is used with the azimuthal measurement number of 36.
(2) Otherwise, if the $MNAM_{audio}$ in Table IV is larger than 36, then the "unknown" HRTFs up to 20 kHz in the comparison subset cannot be reconstructed by interpolation, because the azimuthal measurement number of 36 in pattern 1 is the densest uniform sampling condition that could be extracted from the 72 measured HRTFs. This is the case for KEMAR $HRTF_{original}$ shown in Table

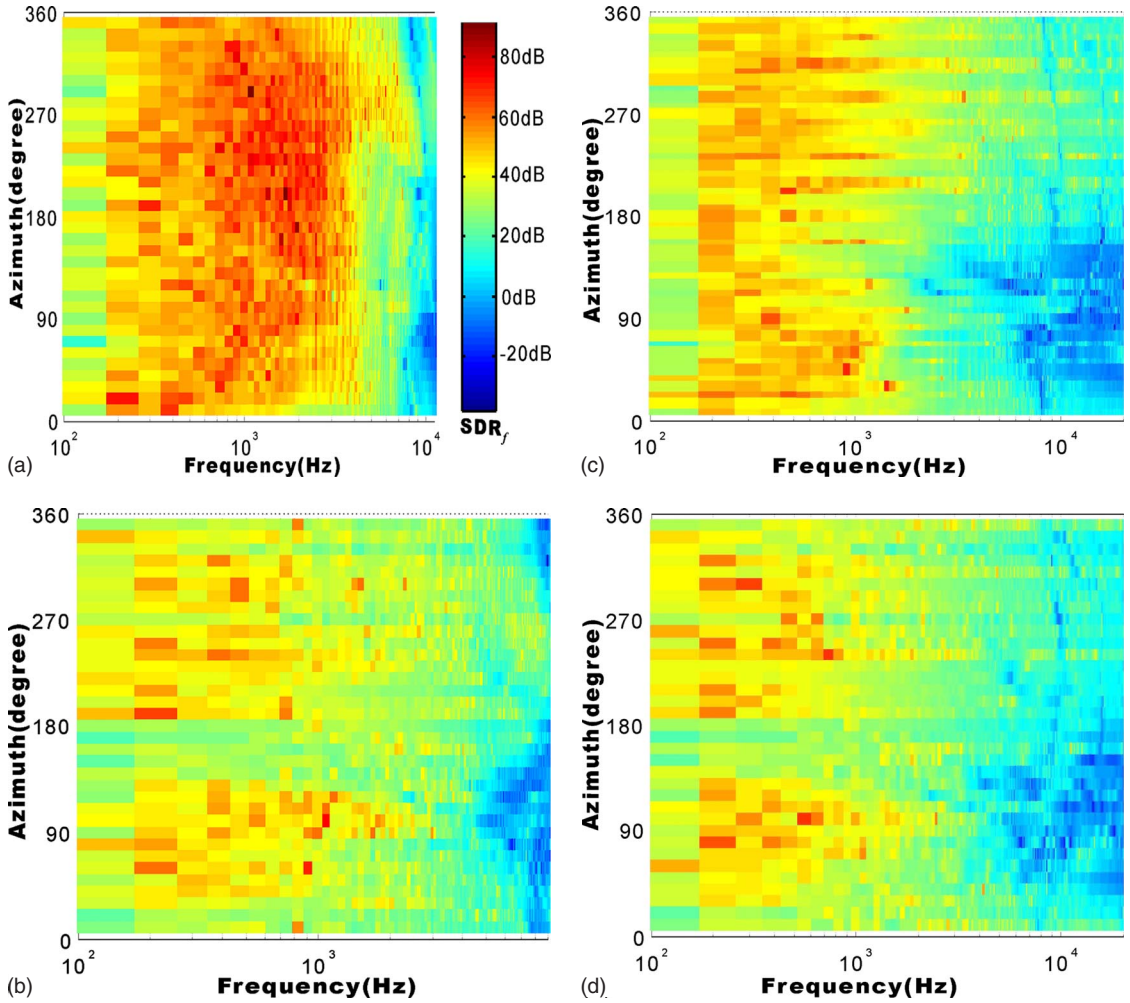


FIG. 4. (Color online) SDR_f varies with frequency and azimuthal position, where (a) is for the $HRTF_{original}$ of KEMAR with the reconstructed frequency up to 11 kHz, (b) is for the $HRTF_{original}$ of subject 23 with the reconstructed frequency up to 9 kHz, (c) is for the $HRTF_{correction}$ of KEMAR with the reconstructed frequency up to 20 kHz, and (d) is for the $HRTF_{correction}$ of subject 23 with the reconstructed frequency up to 20 kHz.

IV. From the MNAMs of KEMAR shown in Fig. 2(a), the MNAM of a frequency range up to 9 kHz is 24, whereas the MNAM of a frequency range up to 11 kHz is 36. Therefore, selecting pattern 1 with the azimuthal measurement number of 36 can yield a wider reconstructed frequency range than pattern 2 with the azimuthal measurement number of 24, according to the azimuthal sampling theorem. So pattern 1 is selected in this case with a reconstructed frequency range up to 11 kHz.

The selected pattern corresponding to each HRTF set and representation, and resulting reconstructed frequency range are shown in rows 4 and 5 of Table IV respectively.

To evaluate the azimuthal interpolation error quantitatively, the signal distortion ratio of each reconstructed frequency point is defined as SDR_f,

$$\mathrm{SDR}_f(\theta,f) = 10\log_{10}\left(\frac{|H(\theta,f)|^2}{|H(\theta,f)-\hat{H}(\theta,f)|^2}\right) \quad (\mathrm{dB}), \quad (12)$$

where $H(\theta,f)$ denotes the measured HRTF, and $\hat{H}(\theta,f)$ denotes the interpolated HRTF from Eq. (6). Since this work focuses on reconstructed HRTFs of a frequency range, SDR is defined as the average signal distortion ratio across the reconstructed frequency range,

$$\mathrm{SDR}(\theta) = 10\log_{10}\left(\frac{\sum_f |H(\theta,f)|^2}{\sum_f |H(\theta,f)-\hat{H}(\theta,f)|^2}\right) \quad (\mathrm{dB}). \quad (13)$$

B. Results

The SDR_f of KEMAR and subject 23 for both $HRTF_{original}$ and $HRTF_{correction}$ are illustrated in Figs. 4(a)–4(d). The general trend is that the SDR_f of each frequency point decreases with increasing frequency; that is, the interpolation error increases with increasing frequency. This result is consistent with other studies. (Minnaar *et al.*, 2005).

The broadband SDRs for the $HRTF_{original}$ of KEMAR and subject 23 are illustrated in Fig. 5(a) as a function of azimuth. In the figure, a common tendency is that the SDR is relatively low at azimuths 0°–180° contralateral to the left ear (the studied ear) and high at azimuths 180°–360° ipsilateral to the left ear. This phenomenon is due mainly to the head shadow effect, which deteriorates the signal-to-noise

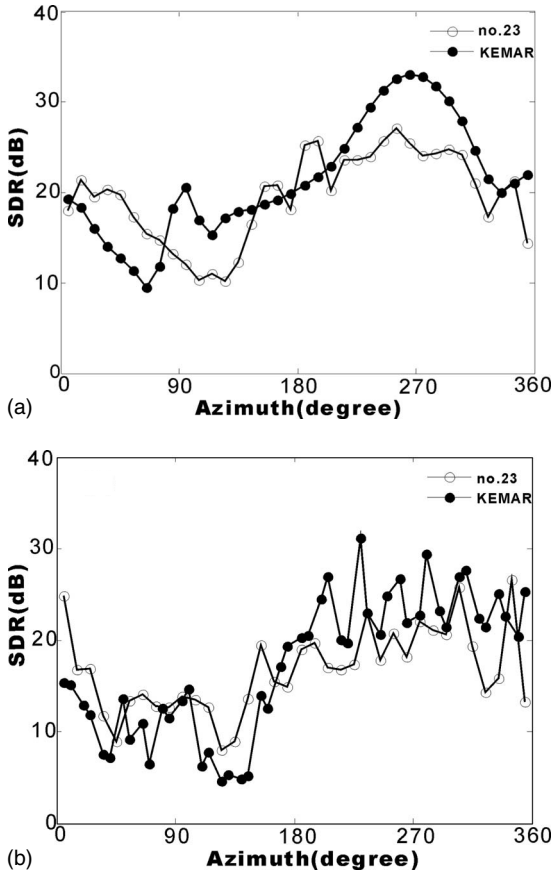


FIG. 5. SDR varies as a function of azimuth in the horizontal plane, with (a) for the $HRTF_{original}$ and (b) for the $HRTF_{correction}$. The filled circles represent the calculation points for KEMAR, and the bare circles represent the calculation points for subject 23.

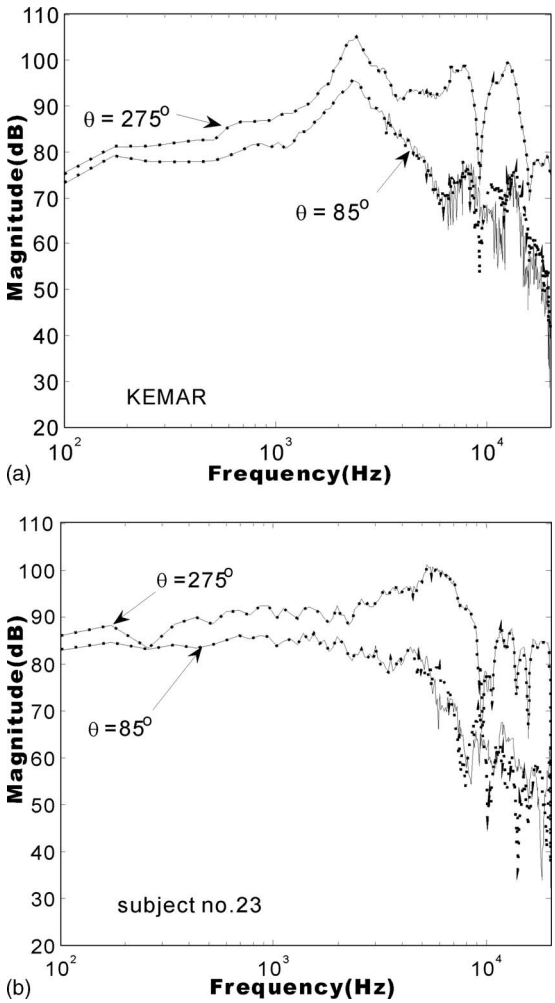


FIG. 6. Comparisons between the measured and interpolated HRTFs under the arrival time correction representation. Solid lines represent the magnitudes of the measured HRTFs and the dashed lines represent the magnitudes of the interpolated HRTFs.

ratio of the contralateral HRTFs. Moreover, the great detail in the contralateral HRTFs, caused by the interference of multi-path waves scattering across the head, would also add to the difficulty of reconstruction. Figure 5(b) is the SDR varying with azimuth in the horizontal plane for the $HRTF_{correction}$ of KEMAR and subject 23. As with the $HRTF_{original}$, the SDR for positions ipsilateral to the left ear is always larger than positions contralateral to the left ear.

The SDR_{mean} (averaging SDRs on azimuth) are shown in row 6 of Table IV. The SDR_{mean} for KEMAR $HRTF_{original}$ is 21.4 dB, and for subject 23 is 19.6 dB. The SDR_{mean}, however, decreases slightly by 3–4 dB with the arrival time correction. This reduction in SDR_{mean} is due to the fact that the interpolation result with the arrival time correction may suffer from the summation of the errors introduced in the process of interpolating the H_τ and the time displacement separately.

To demonstrate the interpolation performance more directly, the similarity between interpolated and measured $HRTF_{correction}$ in magnitude at two typical azimuths (275° and 85° in the horizontal plane) is shown in Fig. 6. In these figures, both for KEMAR and human subject 23, the interpolation performance at azimuth 275° is better than at azimuth 85°, which is consistent with the calculated SDRs in Fig. 5(b) and SDR_f in Figs. 4(c) and 4(d).

V. PSYCHOACOUSTIC EXPERIMENT

A. Method

The aim of this psychoacoustic experiment is to evaluate the validity of the azimuthal sampling theorem and the resulting interpolation formula through a series of auditory discrimination measurements. The task of the listener is to discriminate whether there are some differences between the stimuli filtered by the measured HRTF and that filtered by the interpolated HRTF for any reason perceived. Since the variation in the HRTF with azimuth is most complicated in the horizontal plane, six azimuths in the horizontal plane, i.e., 25°, 55°, 85°, 115°, 145°, and 175°, were chosen as intended positions in the experiment. As shown in row 4 of Table IV, pattern 1, with an azimuthal measurement number of 36, should be used in the interpolation of human HRTFs according to the azimuthal sampling theorem. The azimuthal measurement numbers of 24 and 18 were also used in this experiment with the same conditions, to demonstrate the interpolation performance when the sampling theorem is not satisfied.

Individualized HRTF (HRIR) data for each listener were used in the experiment. For each listener, the HRTFs may be

slightly left-right asymmetric (Zhong and Xie, 2007). Therefore, the interpolated HRTFs of the left and right ear were derived from the measured HRTFs of the left and right ears, respectively. First, a 5 s white noise, used as original mono stimulus, was convolved with the measured or interpolated HRTFs of the two ears, respectively, and then passed through a pair of low-pass filters whose band width was equal to the reconstructed frequency range. Finally, the resulting binaural signals were reproduced through a pair of Sennheiser HD 250 II headphone. Individualized headphone equalization was performed before rending the binaural stimuli to the headphone (Møller *et al.*, 1996).

The experiment was divided into two stages. The first stage was to evaluate the interpolation performance using 36, 24, and 18 azimuthal measurement numbers, respectively, under the original representation. The second stage was identical but under the arrival time correction representation. A three-interval, two-alternative forced-choice testing method was used in the experiment. In each condition [including the intended position, HRTF representation ($HRTF_{original}$ and $HRTF_{correction}$), azimuthal measurement number, and specific listener], there was a stimulus presentation consisting of three segments, with a 1 s interval between neighboring segments. The length of each segment was 5 s, with 1 s fading in at the beginning and 1 s fading out at the end. The first segment was the reference signal "A" filtered with the measured HRTF, and the second and third segments were either reference signal "A" or object signal "B" filtered with the interpolated HRTF with a random order. There were therefore two kinds of stimulus presentations, AAB and ABA. The listeners were asked to judge which segment of the second and the third was different from the first segment. Six listeners from the database with normal hearing participated in the experiment. Each listener heard 216 stimulus presentations in total, 3 kinds of azimuthal measurement number $\times 2$ representations of interpolated HRTF $\times 6$ intended positions $\times 6$ experiments. In a specific condition, there were 36 judgments for an intended position, 6 listeners $\times 6$ experiments.

B. Results

The proportions correct for six intended positions in the horizontal plane are shown in Fig. 7(a) for the $HRTF_{original}$ and Fig. 7(b) for the $HRTF_{correction}$. The 95% confidence limits for chance performance that were calculated from a binomial model are plotted as the two horizontal lines in the figures. For a significance level 0.05, the region bounded by the two solid lines on the graphs corresponds to a chance performance. That is, if the proportion correct is between 0.34 and 0.66, there is no audible difference between the measured and interpolated HRTFs. We note from Fig. 7(a) that the proportion correct of each position increases with the azimuthal measurement number decreasing from 36, 24 to 18 in turn. It is clear that, for all the positions, the proportions correct for the azimuthal measurement numbers of 18 and 24 exceed the upper confidence bound for chance performance except at azimuth 175°, whereas the proportions correct for the azimuthal measurement number of 36 are between the upper and the lower confidence bound for chance performance. In this experiment, since the reconstructed frequency range is determined by the azimuthal measurement number of 36, the azimuthal measurement numbers of 24 and 18 are insufficient to obtain the interpolated HRTFs exactly according to the azimuthal sampling theorem of the HRTF. As a result, only the azimuthal measurement number of 36 can ensure the interpolation performance in audibility for all positions. Similar results are found in Fig. 7(b) for the time arrival correction representation. Consistency between the theory and experimental result validates the azimuthal sampling theorem and the resulting interpolation formula.

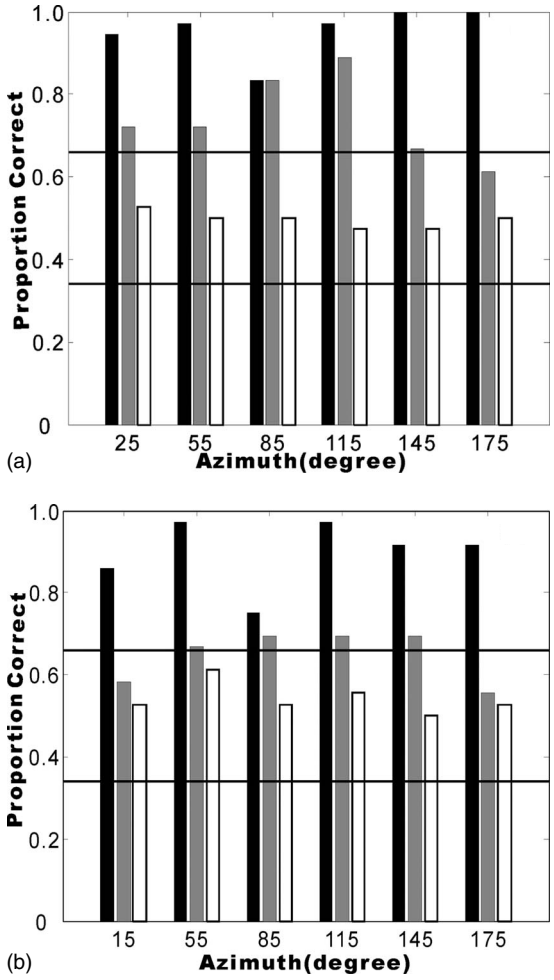


FIG. 7. Bars of proportion correct of judgments for six intended positions in the horizontal plane, with (a) for the $HRTF_{original}$ and (b) for the $HRTF_{correction}$. At each azimuth, the dark bar represents the proportion correct with an azimuthal measurement number of 18, the gray bar represents the proportion correct with an azimuthal measurement number of 24, and the bare bar represents the proportion correct with an azimuthal measurement number of 36. The upper and lower confidence bounds for chance performance are also shown as two horizontal lines.

VI. CONCLUSIONS

Azimuthal Fourier analysis of the HRTF in each elevation plane has been proposed, and an azimuthal sampling theorem for the HRTF and a corresponding interpolation formula deduced. Results show that the azimuthal variation of the HRTF is captured by the azimuthal harmonic coefficients of the HRTF, and if the harmonic expansion of the HRTF is

truncated at order Q according to a certain power criterion, then MNAM for the HRTF is $(2Q+1)$, or equivalently the maximal azimuthal resolution is $360°/(2Q+1)$. If the number of uniform azimuthal measurements in an elevation plane is not less than the MNAM, then the HRTFs at unmeasured azimuths can be reconstructed from measured HRTFs using the interpolation formula proposed above. The azimuthal sampling theorem for the HRTF and the interpolation formula reveal the relations among the spatial variation feature of the HRTF, the number of azimuthal measurements (or the azimuthal resolution), and the interpolation performance, both physically and mathematically.

HRTFs from KEMAR and 52 human subjects under the original and the arrival time correction representation were analyzed. Results show that at a given elevation the MNAM increases with increasing frequency, and at a given sampling point of frequency the MNAM decreases gradually as the elevation plane deviating from the horizontal plane. Moreover, the MNAM for humans is always larger than for KEMAR.

Arrival time correction to HRTF can dramatically reduce $\mathrm{MNAM}_{\mathrm{audio}}$ (the MNAM of the audible frequency range) without introducing obvious mathematical error in reconstruction. As a result, the statistical results of $\mathrm{MNAM}_{\mathrm{audio}}$ with the arrival time correction in HRTFs are proposed as a reference to determine the azimuthal measurement number of HRTF measurement in scientific research.

A psychoacoustic experiment was carried out to evaluate the interpolation performance with and without the arrival time correction in HRTFs. The results show that the interpolation is effective without introducing audible error.

The present results are useful to determine the MNAM needed in arbitrary elevation plane at arbitrary frequency. Only the azimuthal measurement of the HRTF has been investigated, and further work should be carried out on the elevation measurements of the HRTF.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Fund of China Grant No. 10774049 and the Natural Science Fund of Guang Dong Province Grant No. 07300617. The authors wish to express their sincere appreciation for the helpful comments from the reviewers and the associate editor. They also thank Professor Ning Xiang for useful discussions.

Ajdler, T., Faller, C., Sbaiz, L., and Vetterli, M. (**2005**). "Interpolation of head related transfer functions considering acoustics," in the AES 118th Convention, Barcelona, Spain, Preprint 6327.

Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z. (**2002**). "Approximating the head-related transfer function using simple geometric models of the head and torso," J. Acoust. Soc. Am. **112**, 2053–2064.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF database," in Proceedings of the IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, pp. 99–102.

Biscainho, L. W. P., Freeland, F. P., and Diniz, P. S. R. (**2002**). "Using inter-positional transfer functions in 3D sound," in Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. **II**, pp. 1961–1964.

Blauert, J., (**1997**). *Spatial Hearing*, rev. ed. (MIT, Cambridge, MA).

Blauert, J., Brueggen, M., Bronkhorst, A. W., Drullman, R., Reynand, G., Pellieux, L., Krebber, W., and Sottek, R. (**1998**). "The AUDIS catalog of human HRTFs," J. Acoust. Soc. Am. **103**, 3082.

Bovbjerg, B. P., Christensen, F., Minnaar, P., and Chen, X. P. (**2000**). "Measuring the head-related transfer functions of an artificial head with a high directional resolution," presented at the AES 109th Convention, Los Angeles, CA, Preprint 5264.

Burkhard, M. D., and Sachs, R. M. (**1975**). "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**, 214–222.

Evans, M. J., Angus, J. A. S., and Tew, A. I. (**1998**). "Analyzing head-related transfer function measurements using surface spherical harmonics," J. Acoust. Soc. Am. **104**, 2400–2411.

Gardner, W. G., and Martin, K. D. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908.

Genuit, K., and Xiang, N. (**1995**). "Measurements of artificial head transfer functions for auralization and virtual auditory environment," in Proceedings of the 15th ICA, Trondheim, Vol. **II**, pp. 469–472.

Katz, B. F. G. (**2001**). "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements," J. Acoust. Soc. Am. **110**, 2449–2455.

Kulkarni, A., and Colburn, H. S. (**1998**). "Role of spectral detail in sound-source localization," Nature (London) **396**, 747–749.

Matsumoto, M., Yamanaka, S., Tohyama, M., and Nomura, H. (**2004**). "Effect of arrival time correction on the accuracy of binaural impulse response interpolation, interpolation methods of binaural response," J. Audio Eng. Soc. **52**, 56–61.

Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). "Directional sensitivity of sound-pressure levels in the human ear canal," J. Acoust. Soc. Am. **86**, 89–108.

Minnaar, P., Plogsties, J., and Christensen, F. (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis," J. Audio Eng. Soc. **53**, 919–929.

Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**). "Head-related transfer functions of human subjects," J. Audio Eng. Soc. **43**, 300–321.

Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (**1996**). "Binaural technique: Do we need individual recordings?," J. Audio Eng. Soc. **44**, 451–469.

Nishino, T., Kajita, S., Takeda, K., and Itakura, F. (**1999**). "Interpolating head related transfer function in the median plane," in Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, pp. 167–170.

Riederer, K. A. J. (**1998**). "Head-related transfer function measurement," MS thesis, Helsinki University of Technology, Helsinki.

Shinn-Cunningham, B. G. (**1998**). "Applications of virtual auditory displays," in Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 20, 1105–1108.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.

Wenzel, E. M., and Foster, S. H. (**1993**). "Perceptual consequences of interpolating head-related transfer functions during spatial synthesis," in Proceedings of the IEEE 1993 Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 102–105.

Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening. I: Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., Kistler, D. J., and Arruda, M. (**1992**). "Perceptual consequences of engineering compromises in synthesis of virtual auditory objects," J. Acoust. Soc. Am. **92**, 2332.

Xie, B. S. (**2008**). *Head-Related Transfer Function and Virtual Auditory* (National Defense Industry, Beijing, China) (in Chinese).

Xie, B. S., Zhong, X. L., Rao, D., and Liang, Z. Q. (**2007**). "Head-related transfer function database and its analyses," Sci. China, Ser. G **50**, 267–280.

Zhong, X.-L., and Xie, B.-S. (**2005**). "Spatial characteristics of head-related transfer function," Chin. Phys. Lett. **22**, 1166–1169.

Zhong, X.-L., and Xie, B.-S. (**2007**). "Spatial symmetry of head-related transfer function," Chin. J. Acoust. **26**, 73–84.

Specificity of binaural perceptual learning for amplitude modulated tones: A comparison of two training methods

Daniel Kumpik,[a] Jeremy Ting, Robert A. A. Campbell,
Jan W. H. Schnupp, and Andrew J. King
Department of Physiology, Anatomy and Genetics, University of Oxford, Sherrington Building, Parks Road, Oxford OX1 3PT, United Kingdom

(Received 25 April 2008; revised 4 November 2008; accepted 23 January 2009)

The specificity of auditory perceptual learning has been taken as an indicator of the likely locus within the brain at which underlying neuronal changes occur. This study examined interaural level difference (ILD) discrimination learning with sinusoidally amplitude modulated (SAM) tones and whether training-induced threshold improvements generalize from one side of auditory space to the other and to an untrained carrier frequency. A novel, dual-staircase adaptive method was adopted that was designed to prevent participants from identifying the nature of the adaptive track. ILD thresholds obtained with this method were compared with a constant-stimulus technique using otherwise identical stimuli. Adaptive thresholds derived from psychometric functions were found to be biased compared to those obtained from reversals. Although adaptive and constant-stimulus procedures appeared to yield different temporal patterns of learning, no global differences were found between them in terms of training outcomes. These data show that ILD discrimination learning with SAM tones does generalize to an untrained carrier frequency but does not generalize across the midline. This implies that the neural substrate for binaural plasticity is found at a relatively high level of the auditory pathway where information is combined across frequency and where each side of auditory space is represented separately.
 [DOI: 10.1121/1.3082102]

PACS number(s): 43.66.Pn, 43.66.Qp, 43.66.Rq, 43.66.Lj [MW] Pages: 2221–2232

I. INTRODUCTION

The finding that sound localization performance can improve after perceptual training (for a review, see Wright and Zhang, 2006) is highly relevant for hearing impaired populations, for whom binaural cue values may have been altered as a result of either auditory deprivation or stimulation (Boothroyd, 1993; Philibert *et al.*, 2005; Silman *et al.*, 1984). An understanding of the potential extent and generality of perceptual learning is essential for developing optimal assessment and training strategies for these populations and others, including children with linguistic deficits (Constantinides *et al.*, 2003; Kujala *et al.*, 2001; McArthur and Bishop, 2001). To this end, two aspects of perceptual learning—the time-course over which it occurs and the potential for learning to transfer to other stimuli or tasks—are of crucial importance in determining the clinical utility of auditory plasticity.

The degree of transfer of perceptual learning to untrained tasks and stimuli can also provide valuable insights into the possible locus within the brain at which changes in neural circuitry take place and therefore guide subsequent physiological studies. For example, it has frequently been reported that training adult humans to perform visual discrimination tasks leads to improvements that are specific to the location on the visual receptor surface and for the orientation of the trained stimulus (reviewed by Gilbert *et al.*, 2001). This suggests the involvement of early stages in cortical visual processing, where neurons are orientation selective and have small receptive fields that are arranged in a precise, topographic order.

Some of the most dramatic demonstrations of the adaptive capabilities of the adult auditory system have come from studies in which auditory localization cues have been altered (Hofman *et al.*, 1998; Kacelnik *et al.*, 2006; Van Wanrooij and Van Opstal, 2005). However, training can also alter the discrimination of interaural level differences (ILDs) and interaural time differences (ITDs) by normal hearing listeners (Rowan and Lutman, 2007; Spierer *et al.*, 2007; Wright and Fitzgerald, 2001; Zhang and Wright, 2007). Wright and Fitzgerald (2001) found that training-induced improvements in ILD discrimination with pure tones generalized to an untrained standard stimulus location (ILD value), but only at the trained frequency, suggesting that the neural processing of localization cues was altered within frequency-specific channels. On the other hand, generalization across different types of stimuli was reported by Rowan and Lutman (2007), who found that ITD discrimination training with either pure tones or higher-frequency transposed tones led to reduced thresholds for both stimuli and for sinusoidally amplitude modulated (SAM) tones with the same modulation rate and carrier frequency as the transposed tones. This suggests that generalization of ILD learning to other stimulus frequencies might also be found if more complex stimuli, such as SAM tones, are used for training. We examined this question in the

[a]Author to whom correspondence should be addressed. Electronic mail: andrew.king@dpag.ox.ac.uk

present study, as well as the extent to which perceptual learning in a binaural discrimination task is specific for spatial location.

There is a growing body of evidence that intensive training can facilitate improvements in perceptual abilities in humans and animals with central auditory processing or peripheral deficits (Kacelnik *et al.*, 2006; Merzenich *et al.*, 1996; Tallal *et al.*, 1996; Zhou and Merzenich, 2008). Various psychophysical methods have been adopted for the study of perceptual learning. In order to maximize the benefits of training in clinical populations, it is important to compare the degree of learning provided by different methods. This issue seems to have been neglected with respect to transfer of auditory perceptual learning. Since the seminal paper of Levitt (1971), adaptive staircase methods (such as the transformed up-down procedure) have become increasingly widespread in the literature (reviewed by Leek, 2001). Compared with constant-stimulus methods, where the range of stimulus values covered by the underlying psychometric function must be chosen and fixed beforehand, staircase methods require fewer restrictive assumptions and target stimulus values efficiently around a specific threshold on the function based on the listener's previous performance.

The relative flexibility and efficiency of adaptive methods are offset by a number of disadvantages, not least the fact that attentive listeners may become aware of the algorithm being used and try to "second-guess" the strategy, which can lead to distraction from the task at hand, possibly affecting the observed threshold. One remedy that has been suggested to address this particular issue is to use one or more simultaneous staircases, each of which governs the stimulus selection in turn from one trial to the next (e.g., Rammsayer, 1992). Another disadvantage is that adaptive tracking techniques tend to focus rapidly on a predetermined threshold value, which can limit their ability to provide accurate information regarding the slope and shape of the underlying psychometric function. A related problem is the fact that adaptive procedures may not allow an observer sufficient exposure to the range of binaural cue values required for learning to occur (Trahiotis *et al.*, 1990), which could influence the observed time-course of learning or the degree of generalization seen. Finally, several simulation studies have shown that psychometric functions obtained from adaptive data can exhibit an inherent bias that causes the slope to be steeper than that of the underlying psychometric function, again potentially affecting threshold estimates (e.g., Leek *et al.*, 1992; O'Regan and Humbert, 1989; Schlauch and Rose, 1990). This bias appears to be a result of the serial dependency of trials in an adaptive sequence (Kaernbach, 2001) and is exacerbated by other factors, such as the type of staircase, the choice of step size and the number of trials performed (Dai, 1995; Saberi and Green, 1996; Leek *et al.*, 1992). However, obtaining a psychophysical measure of the absolute bias seen in adaptive methods is difficult, unless a full psychometric function is obtained simultaneously. Some studies have reported a "relative" bias between different adaptive methods (e.g., Rowan *et al.*, 2006), whereas others have found that thresholds obtained from adaptive psychometric functions are similar to those obtained from reversals (Amitay *et al.*, 2006) and to those obtained using a constant-stimulus technique (Dai, 1995).

In this study, we compare the relative impact of using a constant-stimulus versus an adaptive training strategy on the learning time-course and generalization pattern seen after ILD discrimination training with SAM tones. We also used these experiments to extend recent findings on the specificity of perceptual learning by examining whether or not training with stimuli containing binaural cues that favor one side of auditory space can enhance performance for stimuli either with a different carrier frequency or presented on the opposite side.

II. METHOD

A. Subjects

A total of 22 listeners took part in the first experiment, in which we compared adaptive and constant-stimulus algorithms for measuring ILD discrimination thresholds, and 31 in the second experiment, which examined the specificity of ILD learning. Seven listeners took part in both experiments, having completed the learning experiment first. All subjects were assessed before testing to ensure normal bilateral audiometric thresholds (i.e., within 20 dB of normal, from 125 Hz to 8 kHz). Ages were in the range 19–42, with a mean age of 24.5 years. All subjects were right-handed. Subjects were recruited through departmental emails and public notices, and received payment for their contribution. All subjects provided informed consent before participation in the study. Ethical approval was provided by the Central University Research Ethics Committee of the University of Oxford.

B. Apparatus

The software package MATLAB (Mathworks, Inc., r.14) was used for stimulus generation, response collection and data organization. Further analyses were conducted in S-PLUS (Insightful Corp, r. 6.2). Stimuli were passed to a Tucker-Davis Technologies (Alachua, FL) RM1 mobile processor for D/A conversion and presented to subjects over headphones (Sony MDR-V500, Tokyo, Japan). Subjects responded by clicking a mouse on one of two response alternatives presented on a flat-screen monitor, which was also used to cue each response period and to provide trial-by-trial feedback. All testing took place in a double-walled sound-attenuating chamber.

C. Stimuli

Prior to testing, head-related transfer functions for 4 and 0.5 kHz pure tones were measured from one of the authors using a similar procedure to that reported previously in the literature (Wightman and Kistler, 1989). From these measurements we estimated the ILD/ITD values associated with 4 and 0.5 kHz tones presented at approximately 5° (for all familiarization runs conducted before testing) and 30° (for the experiment proper) from the midline, and these values

TABLE I. Offset binaural cue values used (± indicates lateralization favoring right or left hemifield).

4 kHz	ILD offset	ITD offset
5°	±1.90 dB	±0.06 ms
30°	±11.42 dB	±0.34 ms
0.5 kHz	ILD offset	ITD offset
5°	±0.70 dB	±0.06 ms
30°	±4.23 dB	±0.36 ms

were used to offset the overall perceived position of stimulus presentation. These offset ITD and ILD values are presented in Table I.

On each trial, two sequential SAM tones, with a carrier frequency of 4 or 0.5 kHz, modulation depth of 100%, and a modulation rate of 80 Hz, were presented binaurally. These stimuli were presented with offset ITDs and ILDs appropriate for sound source directions at either 5° or 30° from the midline. However, while the ITD was held constant throughout the presentation of the stimulus pair, the ILD was varied by a small increment, ΔILD, between the two stimuli. Positive increments (larger ILD on the second stimulus presentation of the pair) gave the percept of the sound source shifting away from the midline, whereas negative increments produced a perceived shift toward the midline. The size of the increment was varied throughout the experiments, either adaptively using the staircase methods described below or chosen randomly from a constant-stimulus set, to determine the subjects' sensitivity to small ΔILDs. We wished to avoid the use of one fixed standard ILD as one of the stimuli in each trial, as this has been viewed as a potential source of the discrimination learning described in other studies (Karmarkar and Buonomano, 2003). The ILD of the first interval stimulus was therefore the offset value plus half the ΔILD and that of the second interval stimulus the offset value minus half the ΔILD, or vice versa, depending on the intended, pseudorandomly determined direction of the ILD shift. Thus, the ΔILDs presented to the subject straddled the offset ILD value appropriate for a sound source located at either approximately 5° or 30° from the midline.

Stimuli were generated digitally, with a sample rate of 48 kHz. Stimulus duration was 400 ms, with an interstimulus interval of 500 ms. The SAM tones were ramped on and off in phase with the stimulus envelope, giving rise/fall times of 6 ms. Stimulus intensity was calibrated such that the maximum sound level (offset ILD+0.5×ΔILD) to be presented to either ear would be 74 dB sound pressure level, and overall presentation level was pseudorandomized at attenuations of 0, −5, and −10 dB to prevent the subjects from using monaural level cues, rather than binaural level difference cues.

D. Procedure

Early pilot work using a single-staircase paradigm highlighted the issue of subjects becoming aware of the adaptive algorithm (Fig. 1), a problem that has been described in some other studies (e.g., Rammsayer, 1992). It appeared that,

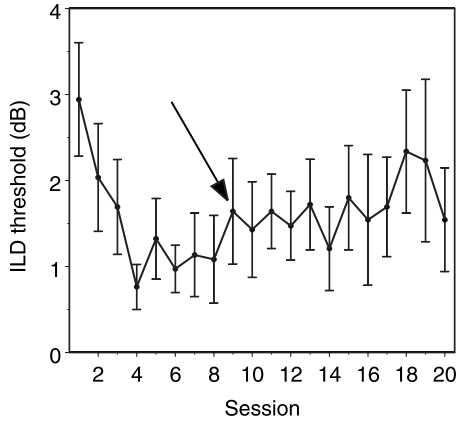


FIG. 1. Example of the confounding nature of awareness of the track in an adaptive staircase method for determining the effects of training on ILD discrimination thresholds. The learning curve represents data from one subject on a single-staircase 3D1U training paradigm. The arrow highlights the first session after the subject became aware of the adaptive algorithm. Errorbars represent one standard error of the reversal points in each session.

even though subjects were capable of performing the task, once they had discovered the algorithm they found it distracting. At times this led to elevated thresholds, because the subjects diverted their attention to trying to second-guess the algorithm, rather than focusing solely on the physical attributes of the stimulus. In the case of the subject represented in Fig. 1, the single-staircase adaptive tracks regularly reversed around stimulus values that were well above the true threshold, based on the performance achieved before the presentation strategy was detected. We therefore devised a similar algorithm that used two staircases running alternately on each consecutive trial. For trial n, stimulus placement was dependent not on the adaptive track pertaining to $n-1$ trials but to $n-2$ trials (see Fig. 2). This allowed for consecutive trials in which the ΔILD value presented was regularly different from one trial to the next and was not dependent on performance on the previous trial, thus reducing the effects of listener expectancy. Each staircase followed a 3-down, 1-up algorithm, with three correct answers leading to a decrease in absolute ΔILD value between the two stimuli and

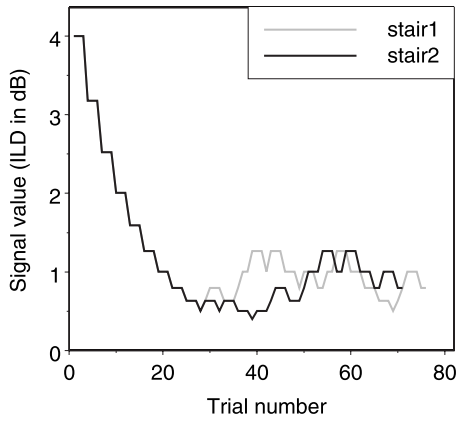


FIG. 2. Example of a dual-staircase run. The first staircase (stair 1) is denoted with a gray line, and the second with a black line (stair 2). In this example, the staircases diverge after around 30 trials.

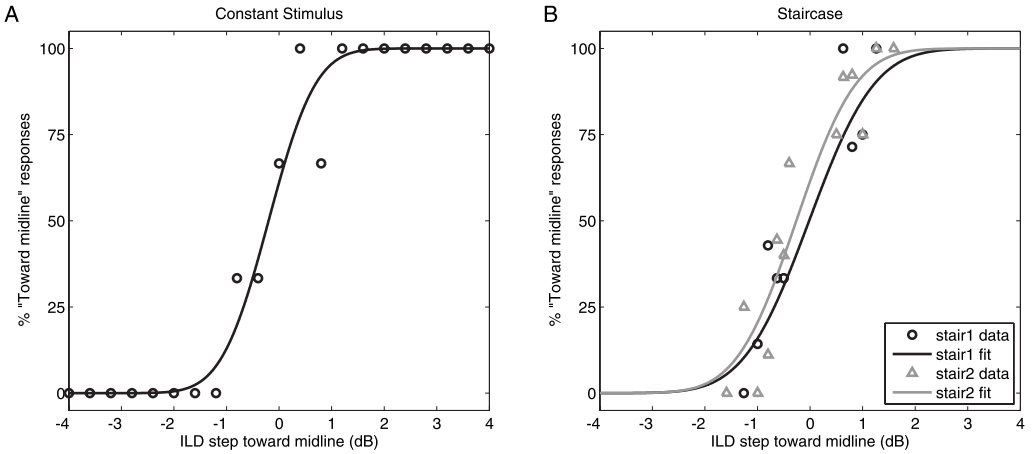


FIG. 3. Examples of psychometric functions fitted using probit analysis for the constant-stimulus method (A) and the dual-staircase method (B). In (B), the black line and circles represent the first staircase, and the gray line and triangles represent the second staircase.

one incorrect answer leading to an increase. For each staircase, this method targeted the 79.4% correct point on the psychometric function (Levitt, 1971).

Pilot data had indicated the presence of serious ceiling effects in untrained subjects (probably caused by the difficulty in performing an ILD discrimination task away from the midline, especially for SAM tones with a 0.5 kHz carrier), and adaptive tracks that frequently reached a ΔILD of 0 dB in highly trained subjects, thus preventing us from obtaining an accurate threshold. We therefore changed the step size logarithmically using a factor of 0.2057 (Saberi, 1995). This procedure results in psychometric functions that are parallel, so that thresholds depend on the performance of the subjects within the range of ILDs tested rather than on the underlying slope of the function. This also prevented the stimulus level from ever reaching zero, reducing the impact of floor effects on the estimated threshold. Logarithmic step sizes were considered to be unnecessary for the constant-stimulus technique, as the selection of stimulus values was predetermined, so that all subjects were exposed to the full range of stimulus values used to plot the psychometric function.

The starting ΔILD for the first run was 4 dB and the starting value for the next run was set two log-steps above the geometric mean of the reversal-derived thresholds from each staircase. A maximum ceiling ΔILD of 10.05 dB was applied, equivalent to four log-steps above the starting ΔILD. Subjects completed 13 reversals per staircase, after which the run was terminated. The constant-stimulus method presented randomized ΔILD values that ranged from 0 to 4 dB in 0.4 dB steps.

The listener's task was identical across experimental stages, groups, and stimulus algorithms. Two SAM tones were presented sequentially on each trial, as described above, differing only by a ΔILD that was determined by the experimental algorithm. In every condition, binaural cue values were displaced from the midline as described above by the introduction of the base ILD and ITD values, upon which the stimulus ILD values were superimposed. Subjects were required to indicate which of the two stimuli had been furthest from the midline.

For experiment 1, both constant-stimulus and adaptive thresholds were extracted by fitting psychometric functions to the data obtained in each run. For constant-stimulus thresholds, trials in which presentation of the most peripheral or lateral stimulus occurred in the first interval were denoted by positive ILD values and those in which it occurred in the second interval by negative values. Psychometric functions were then fitted to the data using probit analysis [Finney, 1971; see Fig. 3(A)]. Because we modeled the data based on the distribution of responses for which the more peripheral stimulus was perceived to have been presented first (i.e., "stimulus shifted toward midline" between the first and second intervals), threshold values equivalent to that yielded by the 3-down, 1-up adaptive method (79.4%) were obtained by taking the mean of the stimulus values yielding 79.4% and 20.6% correct on the fitted function.

For the adaptive method, psychometric functions were fitted to the data from both adaptive tracks in each run [see Fig. 3(B)], and then ILD threshold for each was determined as above. The threshold for a given run was defined as the arithmetic mean of these two values. We also extracted reversal-derived thresholds by taking the geometric mean of the final ten reversals in each staircase and by averaging the values obtained from the two interleaved staircases.

For experiment 2, thresholds were extracted as described above, but for the constant-stimulus data we used all of the data from a given session (six runs) to calculate the daily thresholds. Adaptive thresholds for a given session were defined as the mean of the thresholds obtained from all of the runs performed by the listener during that session.

E. Experimental organization and design

1. Threshold bias in the psychometric function

This experiment was intended to shed light upon the relative integrity of the data obtained from the dual-staircase and constant-stimulus methods, and to provide an indication of the effects of any slope or inherent methodological bias upon measured adaptive thresholds. Subjects performed one run with each method, with base binaural cue values that were offset to 5° left or right of the midline. Half the subjects

performed the constant-stimulus run followed by the adaptive run and vice versa. Because the number of trials in an adaptive run is known to contribute to the bias in the slope of the psychometric function (Kaernbach, 2001; Leek *et al.*, 1992), the number of trials in each constant-stimulus run was made roughly equivalent to the mean number of trials in the adaptive runs, for which the number of trials was variable. This ensured that any bias in thresholds obtained from psychometric functions that was due to differences between the methods could be more accurately assessed, and also ensured that each run lasted a similar length of time. Each constant-stimulus run contained 121 trials, with 11 presentations at each stimulus value. The mean number of trials in each adaptive run was 131.7.

2. Perceptual learning

Listeners were randomly assigned to either the constant-stimulus or dual-staircase algorithm, and also randomly assigned to be either a training subject or a control subject within that algorithm. The experiment was conducted in four phases. All subjects took part in a familiarization phase 2–3 days before the pre-test, designed to acquaint them with the task requirements and the stimuli as far as possible, without inducing binaural learning. To this end, listeners completed eight runs, each comprising 120 ILD discrimination trials in a session lasting 2–3 h, with base binaural cue values that were offset to 5° and a constant ΔILD of 4 dB. This allowed for learning of the task requirements with an easy cue value and at a relatively central position in auditory space, where binaural cue resolution should be well developed (Mills, 1960; Yost, 1974; Yost and Dye, 1988). Runs centered on ILD values corresponding to either 5° left or right of the midline were interleaved throughout this familiarization session, allowing subjects to become accustomed to switching sides during a session in preparation for the pre-test. Subjects were required to score 90% in at least two familiarization runs on each side before progressing to the pre-test.

During the pre-test phase (3–3.5 h), baseline ILD discrimination thresholds were obtained from each participant for each of three conditions: (1) 4 kHz SAM offset 30° to the left, (2) 4 kHz SAM offset 30° to the right, (3) 0.5 kHz SAM offset 30° to the left. Conditions were always run in that order to ensure that order effects were similar across subjects and groups. Before each condition was run, a short (30-trial) practice run was administered, with a 5° offset and ΔILD values of 4 dB.

After the pre-test, training subjects underwent 8 days of ILD discrimination training on condition 1 (4 kHz, 30° left). Training took place over the course of 9–10 days, and lasted for approximately 1 h/day. Control subjects did not participate in the training phase. The post-test phase was identical to the pre-test, and was completed by all subjects 11–14 days after the pre-test.

For the constant-stimulus method, 66 trials were completed in each run, with six presentations at each stimulus level. Six runs were completed in each session at all times. For the dual-staircase method, two runs of 13 reversals per staircase were completed per condition during the pre-test and post-test; during training, four runs were completed per session.

F. Analysis

Any fitted psychometric functions that were over-dispersed were excluded from the analysis (in the case of the bias experiment, this meant that a subject's entire dataset would need to be excluded). Over-dispersion is observed if the variance in a dataset is significantly higher than the variance assumed in the probit model. It can arise when the performance of the subject drifts considerably throughout the course of the run, e.g., as a result of wandering attention or fatigue. We therefore considered over-dispersion as diagnostic of potentially unreliable data, and excluded from further analysis runs for which the value of the deviance for the fitted function was greater than twice the number of degrees of freedom for error. If one of the pair of psychometric functions in an adaptive run was over-dispersed, then that entire run was rejected. Under these criteria, 4 subjects were dropped from the analysis for the bias experiment, leaving 18 subjects to be used in the analysis. Of the remaining data, the degrees of freedom for error for each constant-stimulus function was always 19 and the mean deviance was 20.37 (s.d. 6.74). The mean degrees of freedom for error for the adaptive functions was 11.33 (s.d. 4.17), and the mean deviance was 10.5 (s.d. 4.68).

In the learning experiment, we adopted one further exclusion criterion: a subject's innate inability to perform the task. Subjects whose average pre-test ILD thresholds were greater than two standard deviations above the mean threshold obtained by all (trained and control) listeners within a method were excluded from all analyses. On this basis, two subjects were excluded from the training experiment, leaving 29 subjects for the analysis. Because in the learning experiment the data from all six runs within a session were used to calculate our constant-stimulus thresholds, the data were well-described by the fitted psychometric functions, and exhibited no over-dispersion. The number of degrees of freedom for error was again always 19, and the mean deviance was 20.36 (s.d. 7.21), a value that is extremely close to that obtained in the bias experiment and is therefore indicative of good agreement between the two experiments in terms of goodness-of-fit. The mean standard error of threshold estimate per psychometric function was 0.14 dB in condition 1 (on average, 10.5% of threshold obtained for a given run), 0.33 dB in condition 2 (on average, 10.7% of threshold obtained), and 0.61 dB in condition 3 (on average, 16.5% of threshold obtained). For the adaptive method the total proportion of rejected functions was 9.74%. For the remaining functions the mean degrees of freedom was 10.22 (s.d. 2.9) and the mean deviance was 10.69 (s.d. 4.37). The mean standard error of threshold estimate per psychometric function was 0.38 dB in condition 1 (on average, 25.2% of threshold obtained for a given run), 0.59 dB in condition 2 (on average, 29.4% of threshold obtained), and 0.68 dB in condition 3 (on average, 26.6% of threshold obtained).

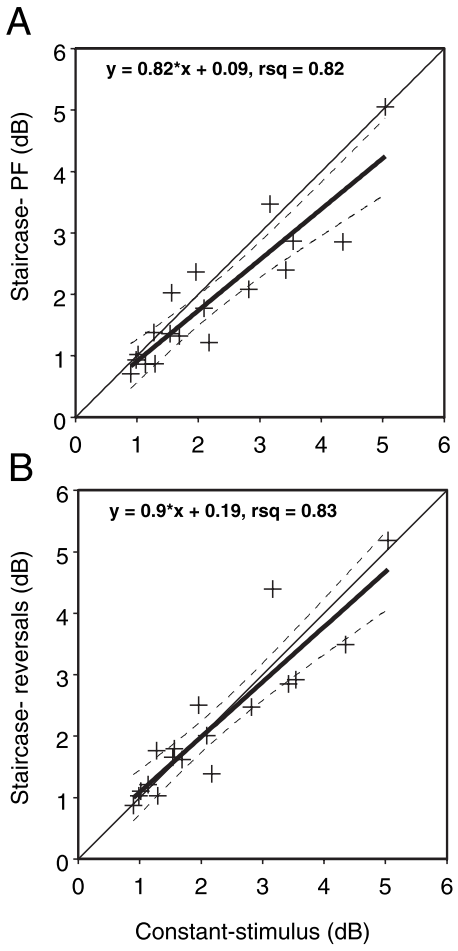


FIG. 4. Bias in the adaptive thresholds derived from psychometric functions (A) and reversals (B) as a function of corresponding constant-stimulus score. The fits are for the non-constrained linear regression models. The crosses indicate thresholds measured for a single subject. Dashed lines indicate 95% confidence intervals for the linear regression fits. In (A), the confidence intervals intersect and are mostly below the unity line, indicating that adaptive thresholds calculated from psychometric functions were generally biased low compared with corresponding constant-stimulus thresholds obtained from the same subjects. In (B), the unity line is completely encompassed by the confidence intervals, indicating that adaptive thresholds were relatively unbiased when calculated from reversals.

The extracted data were then subjected to statistical testing that was appropriate to the analysis in question.

III. RESULTS

A. Assessing the adaptive method

1. Threshold bias in the psychometric function

To assess the amount of bias in the adaptive thresholds, we computed the linear regression of constant-stimulus thresholds against thresholds obtained with the adaptive method (estimated either using reversals or probit fits, as described above). Figure 4 compares the thresholds obtained using the two estimation techniques with the corresponding constant-stimulus thresholds. Adaptive thresholds derived by fitting psychometric functions [Fig. 4(A)] tended to be slightly lower than those obtained from reversals [Fig. 4(B)]. However, for both estimation techniques, regression analysis showed that the intercepts were not significantly different from zero. We therefore constrained the regression intercept term for each to be 0. This yielded r^2 values of 0.96 for each estimation technique, indicating that the relationship between the scores obtained in the constant-stimulus runs and those calculated using either estimation technique from the adaptive data was highly consistent. We then tested the slopes of each of these functions to see if they were significantly different from 1. For the reversal-derived thresholds, this difference was not significant [$t(33)=0.76$, $p=0.45$]. However, the data derived using thresholds calculated from psychometric functions exhibited a slope that was significantly different from 1 [$t(33)=3.21$, $p<0.01$], indicating that, as measured threshold increased, these scores exhibited significant bias compared with the corresponding constant-stimulus scores.

To provide confirmation that the threshold bias described above persisted in the learning data, we also compared the adaptive thresholds derived from psychometric functions with those derived from reversals in the perceptual learning experiment. In the latter case, we collapsed the data across pre-test and post-test for all three experimental conditions. These data provided both a wide range of thresholds and experimental conditions for which this relationship could be compared. Outliers in the learning dataset were identified using a leverage test. If the leverage of any datapoint exceeded $2(p/n)$, where n is the number of observations and p is the number of parameters in the model (i.e., 2), that datapoint was excluded from the analysis (Hoaglin and Welsch, 1978).

Figure 5 compares the psychometric function- and reversal-derived thresholds obtained from both the bias investigation [Fig. 5(A); data derived from Fig. 4] and the learning experiment [Fig. 5(B)]. The regression parameters are almost identical to each other, suggesting that the difference between the two estimation techniques seen in Fig. 4 also applied to the learning data. Neither of the functions shown in Fig. 5 have intercepts that are significantly different from 0, so we constrained the intercepts for both to be the same and fitted models with differing slopes to the data obtained for each. These slopes did not differ significantly [$t(184)=1.88$, $p=0.06$]. These slopes did not differ significantly [$t(184)=1.88$, $p=0.06$]. It can be seen in Fig. 5(B) that the slope of the regression line for the learning dataset shows a small but clear tendency for thresholds derived from psychometric functions to be systematically (mean=8.4%) lower than those obtained from reversals, across the threshold range.

In order to investigate whether or not the psychometric function fitting technique was contributing to the bias observed in these data, we also fitted the pre-test and post-test learning data obtained from the adaptive and constant-stimulus methods using a logit function. Adaptive thresholds calculated using a probit function were, on average, 1.6 times larger than their logit-calculated counterparts, while constant-stimulus thresholds were 1.7 times larger when calculated using the probit function. Paired-sample t-tests showed that both the constant-stimulus thresholds [$t(83)=11.23$, $p<0.0001$] and the adaptive thresholds [$t(168)=13.44$, $p<0.0001$] were significantly lower when calculated using the logit function. To compare how much the choice of fitting technique affected each training method, we

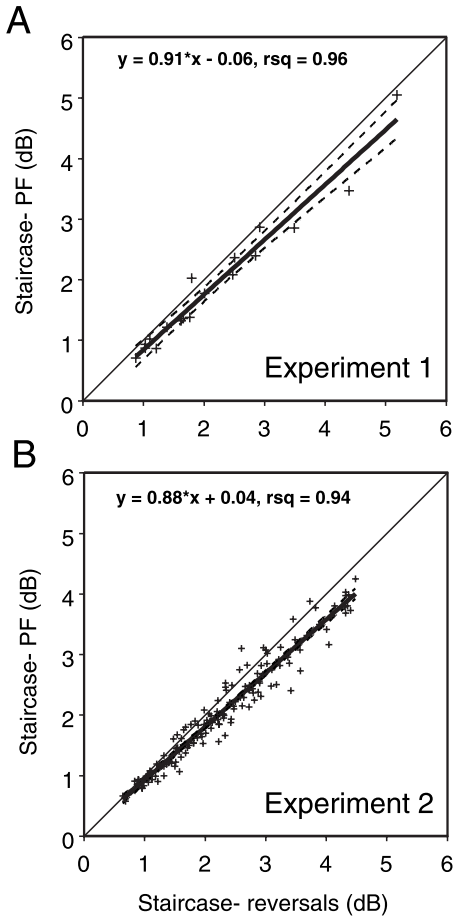


FIG. 5. Comparison of adaptive thresholds calculated from psychometric functions and from reversals for both the bias data (A) and the learning data (B). The dashed lines indicate 95% confidence intervals for the fitted functions; the confidence intervals for the learning data are so close to the function itself that they are only barely visible. For both sets of data, the fitted function lies mostly beneath the unity line, indicating that, when calculated from psychometric functions, the adaptive thresholds from both the bias investigation and the learning experiment exhibited similar degrees of bias compared with reversal-derived thresholds.

calculated difference scores for the thresholds from each fitting technique (probit-derived threshold–logit-derived threshold). A two-sample t-test indicated that the difference scores obtained for each training method were not significantly different from each other [$t(251)=0.42$, $p=0.64$].

These analyses imply that the particular fitting method chosen could significantly affect the amount of bias seen for both constant-stimulus and adaptive thresholds. The logit fits yielded lower thresholds, probably due to the relative "straightness" of the slope of the logit function, indicating that they would match the reversal-derived thresholds even less closely. Equivalent threshold differences were found for the two training methods with probit and logit fitting, suggesting that the choice of fitting method alone could not account for the bias we saw in the adaptive thresholds. Thus, the bias seen in experiment 1 is most likely caused by the adaptive method itself.

Because the reversal-derived thresholds more closely matched the constant-stimulus thresholds obtained in experiment 1 and therefore appeared to be free of bias, the adaptive learning thresholds obtained in experiment 2 were calculated from reversals only.

B. Perceptual learning

Across the training period, the overall mean number of trials for each adaptive staircase within a run was 62.5. The mean number of adaptive trials per subject on each day was 501, compared with the fixed value of 396 for the constant-stimulus method.

1. Time-course of learning in the trained condition

Figures 6(A) and 6(B) show the smoothed individual and group mean ILD learning curves for both methods in the trained condition. Subjects who did not show a numerical improvement in threshold between the pre-test and post-test ($n=1$ for each method) were excluded from further analyses and are not included in this figure. All other subjects showed a gradual, if variable, reduction in ILD thresholds across the training period. In order to characterize the rate of change in improvement for the two methods, we calculated the normalized percentage change in threshold [defined as (pre-test threshold−session×threshold)/pre-test threshold×100] across sessions for all subjects [Figs. 6(C) and 6(D)]. We then fitted linear models, either with or without a quadratic term, to the pooled group mean learning curves using least-squares regression. The constant-stimulus data [Fig. 6(C)] were well fitted by a model with no quadratic term ($r^2=0.9$, $t=8.32$, $p<0.0001$) and when a quadratic term was added it was found not to be significant ($t=1.08$, $p=0.31$), indicating that a linear trend best described the change in thresholds across sessions. The adaptive method, on the other hand, yielded data for which the simplest fit was obtained using a regression model that contained a quadratic term [Fig. 6(D); $r^2=0.9$, $t=2.78$, $p<0.05$].

Because in the adaptive method the number of trials varied from one run to the next, we also fitted the individual normalized learning curves from both techniques against the number of trials completed per session. Individually, the fitted models for the adaptive method were significantly better when they contained a quadratic term in five out of eight cases. In contrast, none of the constant-stimulus learning curves showed a significantly better fit when a quadratic term was included.

These data imply a dissociation between the two methods in terms of the rate of improvement in ILD discrimination performance during training, with the constant-stimulus method exhibiting a linear trend and the adaptive method yielding an improvement that was largely curvilinear. Comparison of Figs. 6(A) and 6(B) shows that the adaptive method generally produced thresholds that were slightly elevated compared to those measured with the constant-stimulus method. However, as can be seen from Figs. 6(C) and 6(D), the overall change in performance between the two methods was very similar.

2. Learning and generalization

The mean thresholds before and after training for each individual subject are shown as pre- vs post-threshold scatter plots in Fig. 7. Trained subjects are plotted in black, untrained controls in gray. Figure 7(A) shows the scores for stimuli with a 4 kHz carrier lateralized to the left, i.e., they

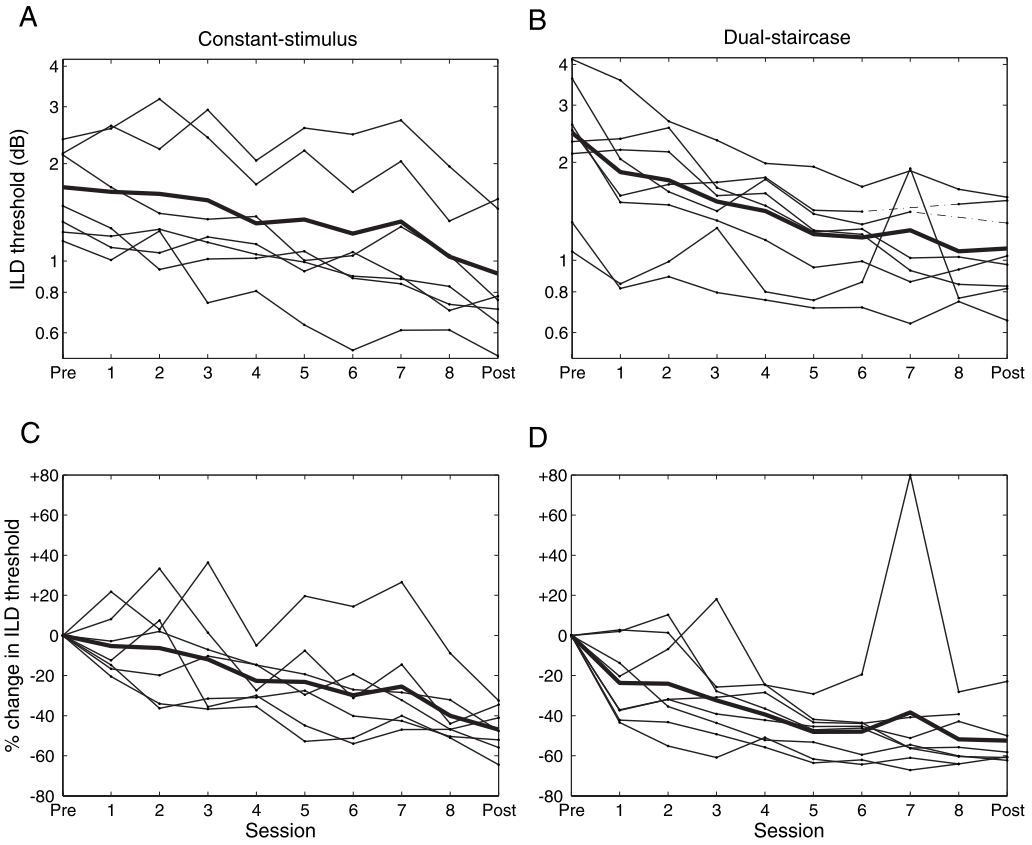


FIG. 6. Mean absolute [(A) and (B)] and normalized percentage change [(C) and (D)] thresholds for each training method as a function of training session. In panel (B) black dashed lines indicate days for which no data were available for a given subject. The thick black line in each plot indicates the group mean for that day. The relative curvature of the mean learning curve for the adaptive data [(B) and (D)] can clearly be seen.

illustrate performance on the trained stimulus condition. It is apparent that the clear majority of trained subjects exhibited improved thresholds after training (all but two of the black symbols lie well below the $x=y$ diagonal). In the "4 kHz right" condition, the datapoints for the trained cohort lie closer to, and are more evenly scattered around, the $x=y$ diagonal [Fig. 7(B)], indicating that any improvement gained by training on the left led to only small or no improvement on the right, i.e., there was little or no transfer to the opposite hemifield. By contrast, in the "0.5 kHz left" condition, we again observed that, in all but one case, the post-test threshold was smaller than the pre-test value [Fig. 7(C)], suggesting that training with a 4 kHz carrier SAM tone had led to improved performance even when tested with an untrained 0.5 kHz carrier SAM tone. In all three conditions, data from the untrained, control subjects are evenly spread around the $x=y$ diagonal, indicating no systematic difference between pre-test and post-test thresholds [Figs. 7(A)–7(C)].

t-tests revealed no significant differences between the methods at pre-test for all listeners in any condition, or for either trained or control listeners at post-test in any condition. In order to increase the power of our analysis, we therefore pooled the data from the two methods for each condition and calculated, for each subject, the normalized percentage threshold change relative to pre-test threshold, again defined as (pre-test threshold − post-test threshold)/pre-test threshold × 100. The observed mean percentage threshold changes were below zero for all groups and all conditions, but only the trained subjects in the "4 kHz left" [$t(16)=7.49$, $p<0.0001$] and the "0.5 kHz left" [$t(16)=3.12$, $p<0.01$] conditions had means that were significantly below zero (confirmed with sign tests; $p=0.002$ and $p=0.0003$, respectively). We saw the same pattern when this analysis was conducted using the absolute change in threshold [$t(16)=5.25$, $p<0.0001$, and $t(16)=4.42$, $p<0.001$ for these two conditions, respectively]. This finding was again confirmed with sign tests ($p=0.002$ for both analyses). Furthermore, t-tests indicated that, for the "4 kHz left" [$t(27)=4.18$, $p<0.001$] and the "0.5 kHz left" [$t(27)=2.2$, $p<0.05$] conditions, the mean percentage change in threshold was significantly greater in the trained cohort than in the untrained controls (confirmed using Wilcoxon rank-sum tests; $p=0.006$ and $p=0.0005$, respectively; see Fig. 8). Again, this pattern of generalization was observed when we analyzed the absolute change in threshold ["4 kHz left": $t(27)=2.81$, $p<0.01$; "0.5 kHz left": $t(27)=3.31$, $p<0.01$], and was confirmed using Wilcoxon rank-sum tests ($p=0.006$ and $p=0.002$, respectively). No significant differences were seen in the "4 kHz right" condition for these analyses.

The mean change in threshold in the trained "4 kHz left" condition was −0.95 dB, or a 46% improvement relative to the mean initial threshold. In the "0.5 kHz left" frequency-transfer condition, we observed a −1.32 dB mean change in threshold, corresponding to a 30% improvement. This change was not significantly different from that seen at "4

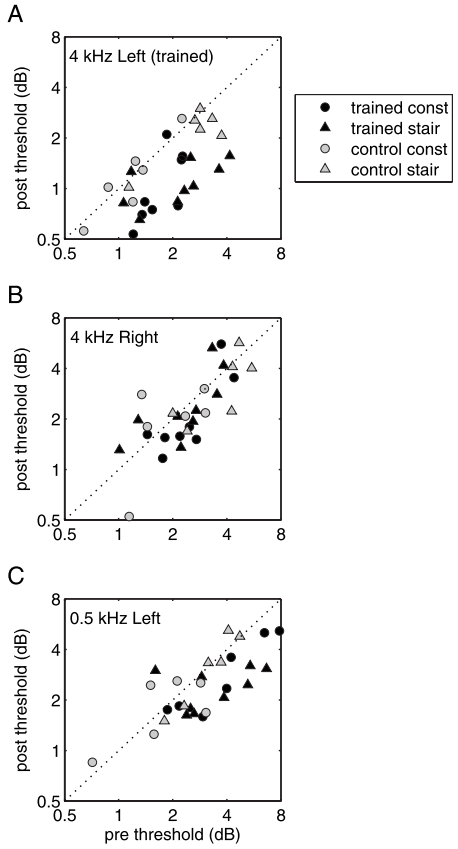


FIG. 7. Pre- and post-test thresholds for all subjects. Results in each of the stimulus conditions (4 or 0.5 kHz carrier, lateralized left or right) are shown in separate panels, as indicated. Subjects trained on the 4 kHz left condition are shown in black, and untrained controls in gray. Subjects trained with constant-level stimuli (const) are shown by the circles, and subjects trained using the adaptive staircase paradigm (stair) are plotted as triangles. In both the trained 4 kHz left and untrained 0.5 kHz left conditions, the trained subjects clearly tend to cluster below the unity line, indicating that post-test thresholds were lower than pre-test thresholds. In contrast, control listeners in all three conditions, as well as trained listeners in the 4 kHz right condition, exhibit thresholds that cluster around the unity line, indicating little change in thresholds between pre-test and post-test.

kHz left" [$t(16)=1.72$, $p=0.11$], indicating that the reduction in thresholds observed with a different carrier frequency was equivalent to that obtained with the trained stimulus.

We also compared the percentage change in threshold between trained and control listeners for each method. The adaptive method yielded significant differences between these groups for the trained "4 kHz left" condition [$t(13)=2.59$, $p<0.05$; confirmed with rank-sum test: $p=0.02$], but not the "0.5 kHz left" condition [$t(13)=1.12$, $p=0.14$]. However, when the absolute change in threshold was used for this analysis, the "4 kHz left" condition approached significance [$t(13)=1.67$, $p=0.06$] and the "0.5 kHz left" condition was significant [$t(13)=2.05$, $p<0.05$; confirmed with rank-sum test: $p=0.03$]. The constant-stimulus percentage change scores were significantly different between the trained and control groups for the "4 kHz left" condition [$t(12)=3.27$, $p<0.01$; confirmed with rank-sum test: $p=0.005$]. In the "0.5 kHz left" condition, a significant difference was also seen between groups [$t(12)=2.08$, $p<0.05$], although this was not confirmed with a rank-sum test ($p=0.1$). When we conducted this analysis using the absolute change in threshold, both the "4 kHz left" condition [$t(12)=3.42$, $p<0.01$; confirmed with rank-sum, $p=0.01$] and the "0.5 kHz left" condition [$t(12)=2.62$, $p<0.05$; confirmed with rank-sum, $p=0.02$] condition yielded significant differences between trained and control listeners. We attribute the apparent contradictions in some of these within-methods analyses to the relatively small sample sizes used in these tests, and the accompanying reduction in statistical power. For both methods, no significant differences were observed between trained and control listeners in the "4 kHz right" condition using any of the above analyses.

Together these results provide strong evidence that training with a 4 kHz SAM tone presented in the left hemifield led to significant improvements in ILD discrimination thresholds, and that these improvements transferred to an untrained carrier frequency presented in the same hemifield, but not to stimuli with the same carrier frequency presented in the opposite hemifield.

IV. DISCUSSION

In this study, we used SAM tones to investigate the time-course and specificity of learning on an ILD discrimination task. Thresholds were estimated using either a standard constant-stimulus method or a novel adaptive task. A methodological investigation was undertaken to compare adaptive thresholds obtained from fitted psychometric functions against those calculated by averaging the reversals in the adaptive track. A comparison was made with thresholds obtained from the same subjects using the constant-stimulus method, and adaptive thresholds were found to be unbiased only when they were calculated from reversals. We also observed differences in the rate of perceptual learning between the two behavioral methods across the training period. However, global learning outcomes in the trained condition were highly comparable in each case. ILD discrimination learning for SAM stimuli whose virtual position was lateralized to the left hemifield of auditory space did not generalize to the opposite hemifield, implying functional segregation of ILD processing for stimuli originating on opposite sides of the midline. However, perceptual transfer of ILD discrimination learning did occur for SAM stimuli presented to the same hemifield used for training, but with a different carrier frequency. Given that ILD learning does not generalize across frequency when pure tones are used (Wright and Fitzgerald, 2001), our results suggest that the amplitude envelope might be a more important cue for ILD discrimination learning than the fine-structure of the stimulus.

A. Methodological assessment

The threshold bias observed when behavioral thresholds were calculated from psychometric functions taken from the adaptive staircase data is consistent with the results of simulations by Leek *et al.* (1992). They found that slope bias in the adaptive psychometric function depends on its underlying true slope, such that, especially for shorter runs, shallower true slopes (and by implication in our analysis, higher thresholds) exhibit more bias than steeper ones (or lower thresholds). This phenomenon is potentially reflected in the

thresholds shown in Fig. 4(A), where it is clear that the higher the constant-stimulus threshold, the more biased is the corresponding adaptive threshold. This is problematic for studies of learning in general, in which the parameters of the psychometric function are likely to change over a period of time and therefore may exhibit more or less bias from one measurement to the next.

Based on the thresholds derived from reversals, the average bias in the adaptive thresholds derived from psychometric functions equated to 8.4% (Fig. 5). Rowan *et al.* (2006) indirectly assessed relative threshold bias for two adaptive methods (one of which was the 3-down, 1-up rule, and both of which are thought to exhibit bias) to be 6.5%. This value is less than that found in our study, as would be expected if the amount by which those adaptive thresholds were biased relative to the underlying psychometric function was within the range observed here.

In contrast to these results, Amitay *et al.* (2006) found little difference between adaptive thresholds obtained from psychometric functions or reversals. Furthermore, Dai (1995) reported that psychometric functions obtained using the constant-stimulus method and both 2-down, 1-up and 3-down, 1-up adaptive methods were equally accurate. However, this conclusion was based upon a comparison between the psychometric functions obtained from each method, and a "total" psychometric function computed by combining the functions derived from all three. If any inherent slope or threshold bias had been present in the adaptive data reported in that study, it may have also been present in the "total" psychometric function.

Simulations carried out by Saberi and Green (1996) indicate that bias in reversal-derived thresholds can be introduced by an inherent imbalance in the relative probabilities of the stimulus level increasing or decreasing at equal distances above or below the target threshold. However, the reversal-derived thresholds in our study showed good agreement with those derived from the constant-stimulus method. In light of this, it seems reasonable to suggest that adaptive thresholds obtained from short runs can be calculated from reversals without compromising the conclusions obtained to any significant degree.

B. Learning in the trained condition

Our results suggest that global learning outcomes for ILD discrimination in the trained condition were relatively unaffected by the choice of psychophysical training method used to induce and measure them. Nevertheless, the learning curves shown in Figs. 6(A) and 6(B) indicate slightly higher thresholds for the adaptive method than for the constant-stimulus method in the earlier training sessions. It is possible that in these sessions the adaptive method was not providing enough exposure to a sufficient range of stimulus levels for maximal learning to occur (e.g., Trahiotis *et al.*, 1990). The relative overall change in performance for the two methods did not, however, differ significantly and post-test thresholds were comparable between them.

The most striking difference between the methods is the temporal pattern of learning shown for each; the rather linear improvement seen for the constant-stimulus method contrasts with the exponential shape of the adaptive learning curve, which shows a large initial drop in threshold but appears to asymptote as training progressed. A similar nonlinear improvement in performance was also observed by Wright and Fitzgerald (2001) when ILD discrimination thresholds were measured using an adaptive staircase procedure. The rapid early improvement could reflect either task-specific or early perceptual learning (e.g., Hawkey *et al.*, 2004) (which seems unlikely, given that substantial familiarization with the task was provided before training began). Because an equivalent initial reduction in thresholds was not observed in the constant-stimulus data, it seems more likely that the overall shape of the adaptive learning curve was influenced by the combined strategies of (1) the logarithmically reducing step size and (2) starting ΔILD values for each run that were taken as two log-steps above the threshold obtained in the previous run. This could have caused large changes in measured threshold at the beginning of training, when the step size between successive stimulus levels was large, and progressively smaller changes, and therefore an asymptotic learning curve, as subjects became more expert with the task and the range of absolute stimulus levels to which they were exposed became smaller.

The thresholds obtained with the constant-stimulus method still appeared to be slowly improving after 8 days of training, suggesting that further ILD learning was possible for the stimuli used in this experiment. If that is the case, then the overall implication is that perceptual learning of ILD discrimination *per se* was relatively linear across time, with small but consistent changes in the degree of learning seen from day to day. Interestingly, the pre-test and post-test thresholds measured in the present study using either training method were lower than those reported by Wright and Fitzgerald (2001), who studied ILD learning with tonal stimuli. Thus, it appears that the use of SAM stimuli may confer an advantage over tonal stimuli for ILD discrimination learning, both in terms of absolute performance thresholds and potentially the degree of learning that is permitted by the underlying neural hardware.

C. Generalization to untrained conditions

Overall, our data do not offer any support for generalization of ILD discrimination learning from one hemifield to the other, implying that perceptual learning for highly lateralized stimuli may utilize perceptual hardware that is functionally separate. This finding is consistent with electrophysiological (e.g., Moore *et al.*, 1984; Brugge *et al.*, 1996; Mrsic-Flogel *et al.*, 2005; Campbell *et al.*, 2006; Harrington *et al.*, 2008) and behavioral studies (Kavanagh and Kelly, 1987; Kelly and Kavanagh, 1994; Malhotra *et al.*, 2004) in animals, and with psychophysical data in humans (Boehnke and Phillips, 1999), showing that each side of auditory space is represented largely contralaterally at higher levels of the auditory system. However, Zatorre and Penhune (2001) examined spatial hearing in human patients with unilateral temporal lobe lesions and found a dissociation in the pattern of localization deficit when the lesion occurred on the right side

of the brain compared with similar damage to the left. They reported that most of the individuals with right auditory cortex lesions were impaired in their ability to perform either a spatial localization or discrimination task in both left and right hemifields. If, as suggested by the results of this study, the right auditory cortex does represent both sides of space, we might expect transfer of ILD discrimination learning across the midline. However, the damaged regions in these patients were relatively large, possibly representing a more extensive region of space than those influenced by sound localization training. Accordingly, our data do not show transfer across the midline, implying that the underlying changes in ILD processing are found at a level of the auditory system where sounds are represented contralaterally.

In contrast to the lack of hemisphere transfer, we found that ILD discrimination learning for high-frequency complex stimuli did generalize to SAM tones with the same modulation rate but a much lower carrier frequency. The same result has been reported by Zhang and Wright (2005) but contrasts with the earlier demonstration by Wright and Fitzgerald (2001) of frequency-specific learning with pure tones. Generalization of perceptual learning is usually interpreted to be indicative of changes at higher loci in the sensory pathway, where information about the trained and generalized stimuli is shared (Gilbert *et al.*, 2001; Gilbert, 1994). Thus, while plasticity of ILD processing for tonal stimuli could potentially occur within any tonotopically-organized structure, beginning with the superior olivary complex where binaural cues are first computed (Yin, 2002), the lack of learning specificity for SAM tones implies the involvement of neurons that integrate information across different frequency channels. This does not mean, however, that binaural learning with more complex sounds, which are lateralized more accurately than pure tones (Bernstein and Trahiotis, 2002), is necessarily restricted to a level of the auditory pathway where frequency selectivity and tonotopicity are lost. Thus, electrophysiological recording studies have shown that neurons in both the inferior colliculus (Biebel and Langner, 2002) and primary auditory cortex (Schulze and Langner, 1999) can be driven by SAM tones with carrier frequencies that lie well outside the excitatory frequency response area defined with tonal stimuli. Although Li *et al.* (2006) found that such stimuli failed to activate inferior colliculus neurons, they did observe responses synchronized to the stimulus envelope when SAM tones with carrier frequencies beyond the excitatory response area were paired with pure tones at the neurons' characteristic frequency. The capacity of these neurons to detect the envelope of the stimulus across a wide range of carrier frequencies is consistent with the transfer of ILD discrimination learning to an untrained carrier frequency (Fig. 8).

Further insights into the neural substrate of auditory perceptual learning will require investigation of the physiological changes that take place as performance improves with training. Although electrophysiological recording studies in animals have so far focused mainly on plasticity of neuronal responses to non-spatial sound attributes, the results of an auditory evoked potential study in humans suggest that short-term training can alter the spatial representation in the auditory cortex (Spierer *et al.*, 2007). While adding to the growing body of evidence that perceptual improvements are associated with cortical plasticity (Dahmen and King, 2007), this does not rule out the possibility that these changes actually originate at a lower level, such as the inferior colliculus.

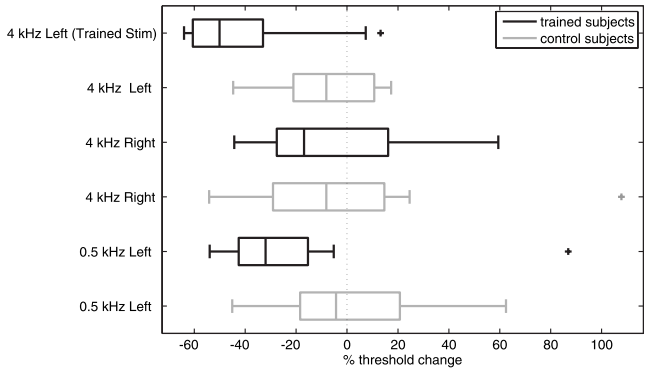


FIG. 8. Box plots showing distribution of threshold changes (post-test minus pre-test threshold expressed as % of pre-test score) for the trained and control cohorts in each of the stimulus configurations. It is clear that only the trained listeners in the 4 kHz left and 0.5 kHz left conditions showed a negative threshold change between pre-test and post-test.

D. Concluding remarks

Provided that thresholds are calculated from reversals in order to avoid adaptive bias, the effects of training on the measured thresholds were comparable for adaptive and constant-stimulus methods for determining ILD discrimination thresholds. Nevertheless, subtle but important differences persisted between the methods that implicate the influence of the adaptive algorithm upon the measured time-course of learning. Finally, the transfer of learning to low frequency ILD discrimination shows that exposure to stimuli with a periodically fluctuating envelope may allow such changes to occur across a wide range of frequencies, whereas the lack of transfer across the midline appears to restrict the site of plasticity to brain regions where spatial cues are represented contralaterally. In both cases, our findings are consistent with the physiological properties of auditory neurons in the midbrain and cortex.

ACKNOWLEDGMENTS

This work was supported by the Wellcome Trust through a Principal Research Fellowship to A.J.K., BBSRC Grant No. BB/D009758/1 to J.W.H.S. and A.J.K., and by the Oxford Centre for Hearing and Balance. We are grateful to John Culling for discussion.

Amitay, S., Irwin, A., Hawkey, D. J. C., Cowan, J. A., and Moore, D. R. (**2006**). "A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naive listeners," J. Acoust. Soc. Am. **119**, 1616–1625.

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.

Biebel, U. W., and Langner, G. (**2002**). "Evidence for interactions across frequency channels in the inferior colliculus of awake chinchilla," Hear. Res. **169**, 151–168.

Boehnke, S. E., and Phillips, D. P. (**1999**). "Azimuthal tuning of human

perceptual channels for sound location," J. Acoust. Soc. Am. **106**, 1948–1955.
Boothroyd, A. (**1993**). "Recovery of speech perception performance after prolonged auditory deprivation: Case study," J. Am. Acad. Audiol. **4**, 331–337.
Brugge, J. F., Reale, R. A., and Hind, J. E. (**1996**). "The structure of spatial receptive fields of neurons in primary auditory cortex of the cat," J. Neurosci. **16**, 4420–4437.
Campbell, R. A. A., Schnupp, J. W. H., Shial, A., and King, A. J. (**2006**). "Binaural-level functions in ferret auditory cortex: Evidence for a continuous distribution of response properties," J. Neurophysiol. **95**, 3742–3755.
Constantinides, H., Rose, M. M., and Moore, D. R. (**2003**). "Training in binaural hearing: towards its use in clinical practice," Int. Congr. Ser. **1254**, 481–485.
Dahmen, J. C., and King, A. J. (**2007**). "Learning to hear: Plasticity of auditory cortical processing," Curr. Opin. Neurobiol. **17**, 456–464.
Dai, H. (**1995**). "On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up-down methods," J. Acoust. Soc. Am. **98**, 3135–3139.
Finney, D. J. (**1971**). *Probit Analysis*, 3rd ed. (Cambridge University Press, Cambridge, UK).
Gilbert, C. D. (**1994**). "Early perceptual learning," Proc. Natl. Acad. Sci. U.S.A. **91**, 1195–1197.
Gilbert, C. D., Sigman, M., and Crist, R. E. (**2001**). "The neural basis of perceptual learning," Neuron **31**, 681–697.
Harrington, I. A., Stecker, G. C., Macpherson, E. A., and Middlebrooks, J. C. (**2008**). "Spatial sensitivity of neurons in the anterior, posterior, and primary fields of cat auditory cortex," Hear. Res. **240**, 22–41.
Hawkey, D. J. C., Amitay, S., and Moore, D. R. (**2004**). "Early and rapid perceptual learning," Nat. Neurosci. **7**, 1055–1056.
Hoaglin, D. C., and Welsch, R. E. (**1978**). "The hat matrix in regression and ANOVA," Am. Stat. **32**, 17–22.
Hofman, P. M., Van Riswick, J. G. A., and Van Opstal, A. J. (**1998**). "Relearning sound localization with new ears," Nat. Neurosci. **1**, 417–421.
Kacelnik, O., Nodal, F. R., Parsons, C. H., and King, A. J. (**2006**). "Training-induced plasticity of auditory localization in adult mammals," PLoS Biol. **4**, e71.
Kaernbach, C. (**2001**). "Slope bias of psychometric function derived from adaptive data," Percept. Psychophys. **63**, 1389–1398.
Karmarkar, U. R., and Buonomano, D. V. (**2003**). "Temporal specificity of perceptual learning in an auditory discrimination task," Learn. Memory **10**, 141–147.
Kavanagh, G. L., and Kelly, J. B. (**1987**). "Contribution of auditory cortex to sound localization by the ferret (Mustela putorius)," J. Neurophysiol. **57**, 1746–1766.
Kelly, J. B., and Kavanagh, G. L. (**1994**). "Sound localization after unilateral lesions of inferior colliculus in the ferret (Mustela putorius)," J. Neurophysiol. **71**, 1078–1087.
Kujala, T., Karma, K., Ceponiene, R., Belitz, S., Turkkila, P., Tervaniemi, M., and Naatanen, R. (**2001**). "Plastic neural changes and reading improvement caused by audiovisual training in reading-impaired children," Proc. Natl. Acad. Sci. U.S.A. **98**, 10509–10514.
Leek, M. R. (**2001**). "Adaptive procedures in psychophysical research," Percept. Psychophys. **63**, 1279–1292.
Leek, M. R., Hanna, T. E., and Marshall, L. (**1992**). "Estimation of psychometric functions from adaptive tracking procedures," Percept. Psychophys. **51**, 247–256.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Li, H., Sabes, J. H., and Sinex, D. G. (**2006**). "Responses of inferior colliculus neurons to SAM tones located in inhibitory response areas," Hear. Res. **220**, 116–125.
Malhotra, S., Hall, A. J., and Lomber, S. G. (**2004**). "Cortical control of sound localization in the cat: Unilateral cooling deactivation of 19 cerebral areas," J. Neurophysiol. **92**, 1625–1643.
McArthur, G. M., and Bishop, D. V. M. (**2001**). "Auditory perceptual processing in people with reading and oral language impairments: Current issues," Dyslexia **7**, 150–170.
Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., and Tallal, P. (**1996**). "Temporal processing deficits of language-learning impaired children ameliorated by training," Science **271**, 77–81.
Mills, A. W. (**1960**). "Lateralisation of high-frequency tones," J. Acoust. Soc. Am. **32**, 132–134.
Moore, D. R., Hutchings, M. E., Addison, P. D., Semple, M. N., and Aitkin, L. M. (**1984**). "Properties of spatial receptive fields in the central nucleus of the cat inferior colliculus. II. Stimulus intensity effects," Hear. Res. **13**, 175–188.
Mrsic-Flogel, T. D., King, A. J., and Schnupp, J. W. H. (**2005**). "Encoding of virtual acoustic space stimuli by neurons in ferret primary auditory cortex," J. Neurophysiol. **93**, 3489–3503.
O'Regan, J. K., and Humbert, R. (**1989**). "Estimating psychometric functions in forced-choice situations: Significant biases found in threshold and slope estimations when small samples are used," Percept. Psychophys. **46**, 434–442.
Philibert, B., Collet, L., Vesson, J. F., and Veuillet, E. (**2005**). "The auditory acclimatization effect in sensorineural hearing-impaired listeners: Evidence for functional plasticity," Hear. Res. **205**, 131–142.
Rammsayer, T. H. (**1992**). "An experimental comparison of the weighted up-down method and the transformed up-down method," Bull. Psychon. Soc. **30**, 425–427.
Rowan, D., Hinton, K., and Mackenzie, E. (**2006**). "Comparison of Levitt- and Zwislocki-type adaptive procedures for stimulus placement in human listeners," J. Acoust. Soc. Am. **119**, 3538–3541.
Rowan, D., and Lutman, M. E. (**2007**). "Learning to discriminate interaural time differences at low and high frequencies," Int. J. Audiol. **46**, 585–594.
Saberi, K. (**1995**). "Some considerations on the use of adaptive methods for estimating interaural-delay thresholds," J. Acoust. Soc. Am. **98**, 1803–1806.
Saberi, K., and Green, D. M. (**1996**). "Adaptive psychophysical procedures and imbalance in the psychometric function," J. Acoust. Soc. Am. **100**, 528–536.
Schlauch, R. S., and Rose, R. M. (**1990**). "Two-, three-, and four-interval forced-choice staircase procedures. Estimator bias and efficiency," J. Acoust. Soc. Am. **88**, 732–740.
Schulze, H., and Langner, G. (**1999**). "Auditory cortical responses to amplitude modulations with spectra above frequency receptive fields: Evidence for wide spectral integration," J. Comp. Physiol. [A] **181**, 651–664.
Silman, S., Gelfand, S. A., and Silverman, C. A. (**1984**). "Late-onset auditory deprivation: Effects of monaural versus binaural hearing aids," J. Acoust. Soc. Am. **76**, 1357–1362.
Spierer, L., Tardif, E., Sperdin, H., Murray, M. M., and Clarke, S. (**2007**). "Learning-induced plasticity in auditory spatial representations revealed by electrical neuroimaging," J. Neurosci. **27**, 5474–5483.
Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., and Merzenich, M. M. (**1996**). "Language comprehension in language-learning impaired children improved with acoustically modified speech," Science **271**, 81–84.
Trahiotis, C., Bernstein, L. R., Buell, T. N., and Spektor, Z. (**1990**). "On the use of adaptive procedures in binaural experiments," J. Acoust. Soc. Am. **87**, 1359–1361.
Van Wanrooij, M. M., and Van Opstal, A. J. (**2005**). "Relearning sound localization with a new ear," J. Neurosci. **25**, 5413–5424.
Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone stimulation of free-field listening. I. Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.
Wright, B. A., and Fitzgerald, M. B. (**2001**). "Different patterns of human discrimination learning for two interaural cues to sound-source location," Proc. Natl. Acad. Sci. U.S.A. **98**, 12307–12312.
Wright, B. A., and Zhang, Y. (**2006**). "A review of learning with normal and altered sound-localization cues in human adults," Int. J. Audiol. **45**, 92–98.
Yin, T. C. T. (**2002**). "Neural mechanisms of encoding localization cues in the auditory brainstem," in *Integrative Functions in the Mammalian Auditory Pathway*, edited by D. Oertel, R. R. Fay, and A. N. Popper (Springer, New York).
Yost, W. A. (**1974**). "Discrimination of interaural phase differences," J. Acoust. Soc. Am. **55**, 1299–1303.
Yost, W. A., and Dye, R. (**1988**). "Discrimination of interaural differences of level as a function of frequency," J. Acoust. Soc. Am. **83**, 1846–1851.
Zatorre, R. J., and Penhune, V. B. (**2001**). "Spatial localization after excision of human auditory cortex," J. Neurosci. **21**, 6321–6328.
Zhang, Y., and Wright, B. A. (**2005**). "Different interaural level difference processing with complex sounds and pure tones," J. Acoust. Soc. Am. **117**, 2562.
Zhang, Y., and Wright, B. A. (**2007**). "Similar patterns of learning and performance variability for human discrimination of interaural time differences at high and low frequencies," J. Acoust. Soc. Am. **121**, 2207–2216.
Zhou, X., and Merzenich, M. M. (**2008**). "Developmentally degraded cortical temporal processing restored by training," Nat. Neurosci. **12**, 26–28.

A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function

Alan Kan,[a] Craig Jin,[b] and André van Schaik[c]
Computing and Audio Research Laboratory, School of Electrical and Information Engineering, University of Sydney, Sydney, New South Wales 2006, Australia

(Received 3 September 2008; revised 6 January 2009; accepted 20 January 2009)

A method for synthesizing near-field head-related transfer functions (HRTFs) from far-field HRTFs measured using an acoustic point-source of sound is presented. Near-field HRTFs are synthesized by applying an analytic function describing the change in the transfer function when the location of a sound source changes from the far-field to the near-field: the distance variation function (DVF). The DVF is calculated from a rigid sphere model and approximates the change in the frequency-dependent interaural level cues as a function of the change in sound source distance. Using a sound localization experiment, the fidelity of the near-field virtual auditory space (VAS) generated using this technique is compared to that obtained by simply adjusting the intensity of the VAS stimulus to simulate changes in distance. Results show improved distance perception for sounds at simulated distances of up to 60 cm using the DVF compared to simple intensity adjustment, while maintaining directional accuracy. The largest improvement for distance perception were for sound sources located to the side and within 40 cm. When intensity was removed as a cue for sound source distance from near-field sounds generated using the DVF, results showed some discrimination of sound source distances but, in general, distance perception accuracy was poor. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3081395]

PACS number(s): 43.66.Qp, 43.66.Pn [JCM] Pages: 2233–2242

I. INTRODUCTION

A head-related transfer function (HRTF) or its time-domain representation, the head-related impulse response (HRIR) describes the change in the sound pressure from a point in the free-field to a listener's ear drum. A separate HRTF exists for each sound source position. Commonly, the HRTF is defined for a planar sound source; however, in this paper we will consider the HRTF for an acoustic point-source of sound. For an acoustic point-source of sound, HRTFs do not vary substantially for distances beyond 1 m since the incoming sound waves are approximately planar, but Brungart and Rabinowitz (1999) showed using acoustic measurements on a manikin that HRTFs can vary substantially for distances within 1 m. This variation is due to the curvature of the wavefront from the point-source of sound and its interaction with the subject's head, torso, and pinnae. Hence, it can be seen that the HRTFs within 1 m are dependent on the distance from the head due to the wavefront characteristics of the sound source.

For an acoustic point-source of sound located off the mid-sagittal plane, Brungart and Rabinowitz (1999) showed that the interaural level difference (ILD) increases as the distance to the head decreases, especially when the sound source is located within 0.5 m from the head. In particular, the low-frequency ILD (<500 Hz), which is usually small at sound source distances greater than 1 m, can be on the order of 20 dB for a sound source located at 0.12 m. Also, the HRTF for the contralateral ear shows increasing high-frequency attenuation with decreasing sound source distance. Results from human sound localization experiments conducted by Brungart *et al.* (1999) and Brungart (1999) suggest that the level differences in the HRTFs are a salient cue for distance localization for sound source distances within 1 m. From acoustic manikin measurements, Brungart and Rabinowitz (1999) also showed the existence of a shift in the spectral features of the HRTF, particularly at the high frequencies, as a function of distance, which is due to an "acoustic parallax" effect. This occurs when the angle between a sound source and the ear changes substantially when the sound source is moved closer to the head. The measurements also exhibit a small increase in the interaural time difference (ITD) at very close sound source distances (<0.25 m). However, Brungart and Rabinowitz (1999) argued that the small increase in ITD is likely to be insignificant and that the ILD in this region is the more dominant cue for sound source distance localization. To distinguish the region where the HRTFs exhibit significant variations with distance from the region where the HRTFs are relatively constant, the terms "near-field" and "far-field" will be used in this paper to refer to these two regions, respectively. A distance of 1 m is taken as the boundary between the two regions.

For the generation of virtual auditory space (VAS), it is common practice to acoustically measure HRIRs for a modest number of different sound source directions at a fixed radius of 1 to 2 m from the center of the listener's head. To estimate an HRIR for a sound source direction that has not been measured, numerous interpolation techniques have been proposed [for example, see Wenzel and Foster (1993); Har-

[a]Electronic mail: akan@ee.usyd.edu.du
[b]Electronic mail: craig@ee.usyd.edu.au
[c]Electronic mail: andre@ee.usyd.edu.au

tung *et al.* (1999); Carlile *et al.* (2000)]. VAS is generated by filtering sounds with HRIRs and presenting the filtered sounds over headphones [see Carlile (1996)]. Near-field VAS requires near-field HRIRs. However, Brungart (2002) highlighted a number of technical difficulties in recording near-field HRIRs on a human subject. First, it is difficult to obtain a small, broadband acoustic point-source which is necessary since the sound source needs to remain compact at relatively close measurement distances and to avoid sound reflections from the loudspeaker back into the microphones. Second, the subject's head position needs to be tightly constrained so that small head movements do not cause substantial angular errors in the recorded HRTFs. Third, an increased number of HRIR measurement points are needed to cover the variation in the HRTFs at the different distances which prolongs the measurement process and can be uncomfortable for the subject. Due to these technical difficulties, it has so far been difficult to obtain HRTFs measured on human subjects to accurately synthesize near-field sound sources in virtual auditory displays.

In this work, we present a new method for generating near-field VAS which involves synthesizing near-field HRIRs from far-field HRIRs. The near-field HRIRs are synthesized using analytical results obtained from a rigid sphere model. A rigid sphere is a reasonable first-order approximation of the human head and allows exact calculation of the pressure at the surface of the rigid sphere due to a stationary sound source in any direction and distance, at any frequency. By calculating, for a particular direction, the ratio of the pressure at the surface of the rigid sphere due to a far-field sound source to the pressure on the rigid sphere due to a near-field sound source for all frequencies of interest, a function can be obtained that describes the change in the transfer function of a sound from the far-field distance to the near-field distance. We call this function the distance variation function or DVF. A near-field HRTF is synthesized by combining the far-field HRTF, corresponding to the same direction, with the calculated DVF. The main advantage of the DVF method for synthesizing near-field VAS is that a near-field HRTF can be analytically calculated for any distance given a far-field HRTF. Since no near-field HRIRs need to be recorded, the technical difficulties with recording near-field HRIRs on a human subject are avoided.

A sound localization experiment was conducted to test the perceptual fidelity of the VAS generated using the HRTFs synthesized by the DVF method. The same experiment was also conducted with near-field sounds synthesized by simply adjusting the intensity of the VAS stimulus. The results of the experiments are presented and compared. In a further experiment, the intensity of the near-field sounds synthesized using the DVF method was held constant to investigate whether the remaining cues introduced by the DVF are salient for sound source distance perception.

II. METHODS

A. Calculation of the DVF

The DVF is calculated using a rigid sphere model. The function describing the pressure of a sinusoidal point-source of sound from any direction and distance in space to the surface of a rigid sphere was derived by Rabinowitz *et al.* (1993) and has been experimentally verified on a bowling ball by Duda and Martens (1998). The pressure at a point X on the surface of a rigid sphere of radius a due to a sinusoidal point-source of sound of angular frequency ω, at a distance r away from the center of the sphere, is given by

$$p(a,\omega,\theta,r) = -kr\sum_{m=0}^{\infty}(2m+1)\frac{h_m(kr)}{h'_m(ka)}P_m(\cos\theta)e^{-ikr}, \quad (1)$$

where $h_m(kr)$ is the spherical Hankel function of the first kind of order m and $h'_m(ka)$ is its first derivative at radius a, $k=\omega/c$ is the wavenumber, c is the speed of sound in air, $P_m(\lambda)$ is the Legendre polynomial of degree m, and θ is the angle between a vector from the center of the sphere to the point X on the surface of the sphere and a vector from the center of the sphere to the sound source. By evaluating Eq. (1) for all frequencies of interest, the transfer function at point X can be obtained for a particular source location. In our experiments, we calculated the transfer function for 256 frequency bins up to the Nyquist frequency.

To synthesize a near-field HRTF, the DVF for a sound source at a near-field distance d_n is calculated as the ratio of the pressure arising from a sound source at distance d_n to the pressure arising from a sound source at a distance d_f, where d_f is the distance at which the far-field HRTF was measured. In other words, we calculate

$$\mathrm{DVF} = \frac{p_n(a,\omega,\theta,d_n)}{p_f(a,\omega,\theta,d_f)} \quad (2)$$

using Eq. (1), for all frequencies of interest, where a is the radius of the listener's head and can be determined through measurement. In our experiment, the radius of the listener's head was estimated using the ITD model proposed by Kuhn (1977):

$$\mathrm{ITD} \approx \frac{3a}{c}\sin\theta_{\mathrm{inc}}, \quad (3)$$

where θ_{inc} is the angle of incidence of the sound source relative to the mid-sagittal axis. ITD values were obtained by cross-correlation of the measured left and right HRIRs below 5 kHz. To determine the angle θ to be used in Eq. (2), the positions of the ears were assumed to be at an angle of 100° on either side of the mid-sagittal plane on the audio-visual horizon for all subjects, in line with Duda and Martens (1998) and Blauert (1997). A separate DVF is calculated for each ear. To synthesize a near-field HRTF from a far-field HRTF, we calculate

$$\mathrm{HRTF}(d_n) = \mathrm{DVF} \times \mathrm{HRTF}(d_f). \quad (4)$$

To synthesize sounds at multiple locations and distances, a DVF needs to be calculated for every sound source distance and direction and for each ear.

B. HRIR measurement

The far-field HRIRs used in this experiment were recorded individually for each subject in an anechoic chamber

equipped with a loudspeaker (VIFA-D26TG-35) mounted on a robotic arm. The robotic arm can accurately position the loudspeaker to within a fraction of a degree, at any point on the surface of an imaginary sphere of 1 m radius except for positions greater than 45° below the audio-visual horizon. The subject was seated with his head in the center of the measurement sphere and aligned to the axes of the measurement system with the aid of a laser-alignment system. A chin-rest was provided and an electromagnetic sensor mounted on a headband worn by the subject monitored his head orientation during the measurement process. A light emitting diode (LED) display provided feedback to the subject with regard to the alignment of his head.

HRIRs were recorded using a "blocked-ear canal" technique, as described by Møller (1992), by placing miniature microphones (AuSim) at the entrance of the ear canal in both ears. Complementary Golay codes were used as stimuli and the HRIRs were obtained using the steps described by Zhou *et al.* (1992). In order to improve the signal-to-noise ratio, 16 repetitions of the Golay codes, each with a 1024 sample length, were recorded at each position. Tucker Davis Technology (TDT) system II hardware, interfaced with customized MATLAB software, was used to play and record the codes at an 80 kHz sampling rate. The stimulus duration at each recording position was just under 1 s. HRIRs were recorded for 393 different sound source directions, upwards of 45° below the audio-visual horizon and equally distributed around the subject's head.

After recording, a number of post-processing steps were applied to the HRIRs. First, the HRIRs were low-pass filtered at 16 kHz and resampled at 48 kHz. Second, since the HRIRs recorded with our system are at the limits of the noise floor for frequencies below 500 Hz, the recorded HRIRs were compensated below 500 Hz for each direction according to the frequency response derived from the rigid sphere model in Eq. (1) by replacing the HRTF magnitudes for frequencies below 500 Hz with those calculated by Eq. (1) and scaled relative to the level of the measured HRTF at 500 Hz. It should be noted that the HRIRs used in this experiment contain the transfer function of the measurement loudspeaker. Third, the magnitude response of the ear canal resonance up to 4 kHz, measured on a Brüel and Kjær Head and Torso Simulator (HATS) manikin (Type 4128), was also added to the HRIRs to compensate for the missing ear canal resonance caused by the blocked-ear recording and the use of in-ear tube phones for VAS presentation. The magnitude of the ear canal resonance was obtained from a blocked-ear recording made with the Ausim recording microphones on the HATS manikin and an open-ear recording made with the ear simulator microphones inside the HATS manikin. Assuming both the Ausim microphones and ear simulator microphones have a flat frequency response, the ear canal resonance can be obtained by subtracting the blocked-ear canal HRTF from the open-ear canal HRTF in the log-magnitude domain. The ear canal resonance obtained has a similar response as that presented in Hammershøi and Møller (1996) for the mean ratio of the pressure at the blocked-ear to the pressure at the ear-drum for 12 subjects, and was applied to the recorded far-field HRIRs as a minimum-phase filter.

C. Stimulus generation

A psychoacoustic experiment was conducted to test the perceptual fidelity of the VAS generated by the synthesized near-field sounds. White noise bursts, 400 ms in duration with 5 ms raised-cosine onset and offset ramps, were used as stimuli for the experiment. A different white noise burst was generated for each sound source location. In this experiment three conditions were tested, in which the near-field stimuli were generated by (1) filtering white noise bursts with near-field HRIRs synthesized by the DVF, henceforth referred to as the DVF-generated stimulus condition; (2) filtering white noise bursts with far-field HRIRs and scaling the intensity of the stimulus inversely proportional to the simulated sound source distance, henceforth referred to as the intensity-scaled stimulus condition; and (3) filtering white noise bursts with near-field HRIRs synthesized by the DVF but with constant intensity, henceforth referred to as the constant-intensity DVF-generated stimulus condition.

To generate stimuli for the DVF-generated stimulus condition, white noise bursts were convolved with near-field HRIRs synthesized by the DVF according to Eq. (2) and (4). A constant post-scaling factor was then applied to the stimuli to ensure that the stimuli at positions closest to the ears would be at a comfortable listening level.

In the intensity-scaled stimulus condition, we desired a scaling factor that would change the intensity of the stimuli with respect to the simulated distance and account for the fact that the distance between a sound source and the two ears can be substantially different when the source is located in the near-field and close to the interaural axis. For a real sound source, the intensity of the sound is inversely proportional to the distance, that is $I \propto 1/d$. To account for the different distances to the two ears, the intensity can be assumed to be approximately proportional to the mean of the two distances, that is,

$$I \propto \frac{1}{2}\left[\frac{1}{d_l} + \frac{1}{d_r}\right], \tag{5}$$

where I is the intensity and d_l and d_r are the distances between the sound source and the left and right ears, respectively. The scaling factor, S, can be defined as the ratio of the intensity level for a sound source at a near-field distance over the intensity level for a sound source at 1 m. Assuming the distance to the two ears are identical for a sound source at 1 m (e.g., along the mid-line), then

$$S = \frac{I_{\text{near field}}}{I_{1\text{ m}}} = \frac{50}{d_l} + \frac{50}{d_r}, \tag{6}$$

where d_l and d_r are in centimeters. The stimuli for the intensity-scaled stimulus condition were thus obtained by convolving white noise bursts with far-field HRIRs and scaling by the numerical factor S. An additional post-scaling factor was applied to the stimuli to ensure that a sound directly in front of the subject at a distance of 1 m would have the same root-mean-squared energy as a sound for the DVF-generated stimulus condition for the same position.

In the constant-intensity DVF-generated stimulus condition, near-field sounds were generated as described for the

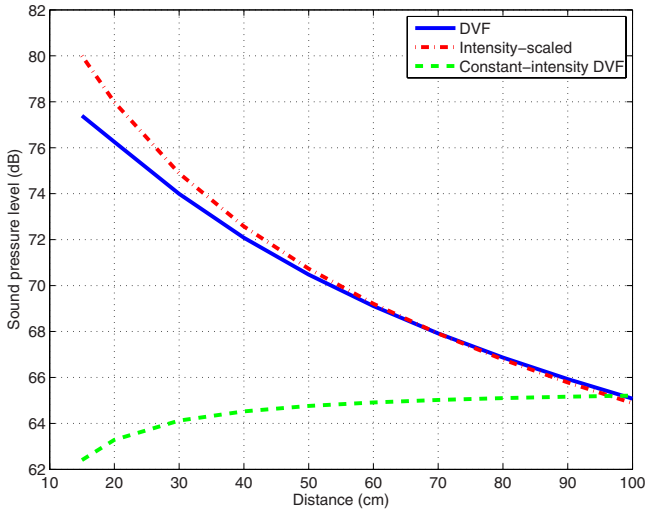


FIG. 1. (Color online) The stimulus level presented to the right ear for a sound source located directly in front of a subject at different simulated distances. Similar stimulus levels were presented to the left ear for the same sound source positions.

DVF-generated stimulus condition with an additional scaling factor of $1/S$ [refer to Eq. (6)] applied to the sounds to achieve approximately constant intensity for all simulated sound source distances. This condition was included to investigate the salience of the remaining cues introduced by the DVF for sound source distance perception when the intensity cue is absent.

Figure 1 shows the range of stimulus levels at the left and right ears for a sound source located directly in front of the subject at different simulated distances for the different conditions.

Near-field sounds were generated for 76 of the 393 HRIR measurement directions, in each of the following four distance ranges (in cm): [10–20], [25–35], [40–60], and [60–100]. Within each range, distances were normally distributed around the mean of each range. The 76 directions were equally spaced around the head between −40° and 40° elevation from the audio-visual horizon.

D. Experimental setup

A sound localization task was used for the experiment whereby the subjects indicated the perceived location of the near-field sounds by placing an electromagnetically-tracked sensor, mounted on the tip of a wand, at the perceived location. Brungart *et al.* (2000) found this method to be the least biased and more accurate than other types of response methods for indicating the perceived location of a sound. An electromagnetic tracking system (Polhemus Fastrak), with two sensors, was used to measure the perceived sound source location relative to the center of the listener's head. One of the sensors was mounted on the tip of the wand, and the other on a rigid headband worn by the listener. The transmitter of the electromagnetic tracking system was located about 50 cm behind the subject's head on a wooden pole, as shown in Fig. 2, and connected to a computer controlling the experiment from an adjacent room. A handheld pushbutton was given to the subject to trigger the computer and an LED display was used to provide head-orientation feedback to the listener prior to each stimulus presentation. Sounds were presented through in-ear tube phones (Etymotic Research ER-2) and played through an RME HDSP soundcard connected to the computer. The experiments were conducted in an ordinary room with no acoustical treatment. The room was lit during the experiment so that the subject could see the exact placement of the sensor at the end of the wand.

At the start of each trial the position of the center of the subject's head relative to the sensor on the headband was estimated by sequentially placing the sensor on the wand on the left and right ears and recording the sensor positions. The center of the head was then calculated as the midpoint between the recorded locations of the left and right ears. Prior to each stimulus presentation, the subject aligned his head to a prescribed starting position with the aid of an LED display [Fig. 2(a)]. The subject then pressed the handheld pushbutton causing the position of the sensor on the headband worn on the subject's head to be recorded by the computer to determine the current location of the center of the head and the stimulus to be played. The subject then responded by placing

FIG. 2. (Color online) The sound localization experiment setup. (a) shows the listener at the prescribed starting position and (b) shows the listener indicating the perceived sound source location

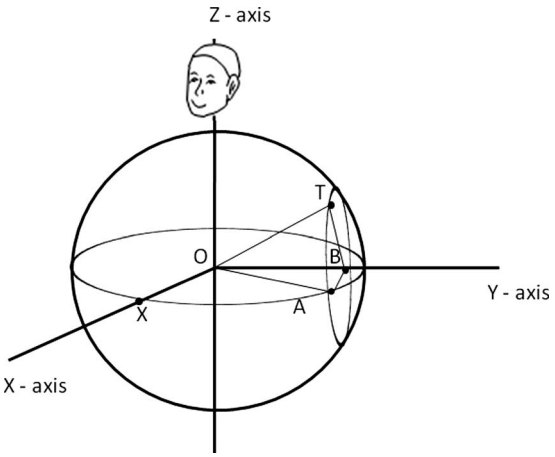


FIG. 3. The lateral-polar angle coordinate system. The lateral and polar angle related to point T are given by $\angle AOX$ and $\angle TBA$, respectively.

the sensor at the end of the wand at the perceived location, as shown in Fig. 2(b), and pressed the handheld pushbutton again. A reading was then taken from the sensor at the end of the wand to capture the perceived location. The perceived direction and distance of the sound stimulus were calculated relative to the location of the center of the subject's head prior to stimulus presentation. The subject then returned to the prescribed starting position for the next stimulus presentation.

In each test session, subjects would localize near-field sounds at each of the four distance ranges for 76 different directions. The stimuli for all directions and distances were presented in random order in each session. Subjects completed five sessions each for each stimulus condition described in Sec. II C. Since the different stimulus conditions were tested in separate sessions, subjects were aware of the method that had been applied to produce the stimuli during the test session. A total of five subjects participated in these experiments. The subjects were all male, between the ages of 26 and 37, and had normal hearing in both ears. All subjects had prior training and experience in human auditory localization experiments involving a head-pointing paradigm [see Carlile *et al.* (1997)] but no new training was given to the subjects for this new localization paradigm.

E. Analysis of results

The results of the psychoacoustic experiments will be presented in terms of direction and distance separately. The directional localization data will be presented using a lateral-polar angle coordinate system [see Fig. 3]. The lateral angle for a particular point T is defined as the horizontal angle away from the XZ-plane given by $\angle AOX$ in Fig. 3. Negative lateral angles (down to −90°) and positive lateral angles (up to 90°) define the left and right hemispheres, respectively. The polar angle corresponding to the point T is given by $\angle TBA$ in Fig. 3. Polar angles are defined with 0° representing the front, 90° representing directly above, 180° representing the back, and 270° representing directly below. The distance localization data will be presented with reference to the center of the head. Scatter plots will be used to display the data in terms of lateral angle, polar angle, and distance localization, where the target angle or distance is plotted against the response angle or distance. Perfect localization, that is, when the response location matches the target location, is indicated by circles along the diagonal from the bottom left to top right corner of the scatter plot. Mean differences between target and response locations across the localization data for all subjects will also be presented along with results from statistical significance tests. Kruskal–Wallis non-parametric analysis of variance (ANOVAs) and *post hoc* analyses (Tukey HSD) were conducted on the data within each condition, as well as across the conditions. An alpha value of 0.05 was used for significance testing. The corresponding non-parametric effect size was estimated by calculating Cliff's d, as described by Cliff (1993), where d estimates the fraction of the total number of samples where the samples in condition A are greater than the samples in condition B, minus the fraction of the total number of samples in condition B that are greater than the samples in condition A. That is, we calculate

$$d = \frac{1}{mn} \sum_i \sum_j d_{ij}, \tag{7}$$

where m and n are the number of samples in conditions A and B, respectively, and

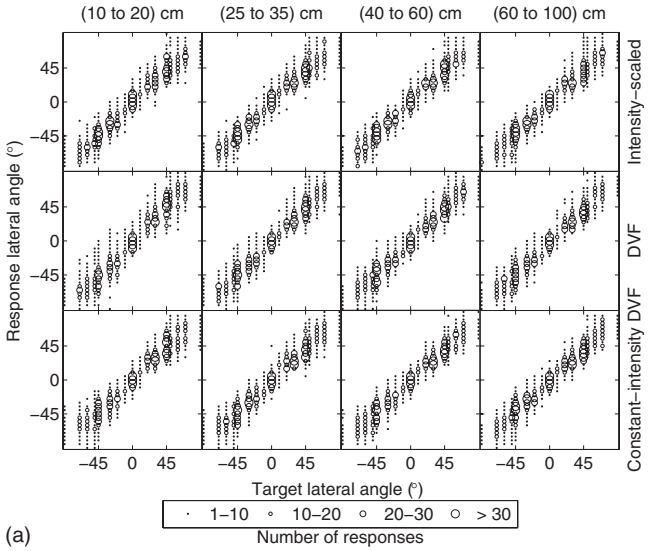
$$d_{ij} = \text{sign}(x_i - x_j), \tag{8}$$

where $i=1,2,\ldots,m$ and $j=1,2,\ldots,n$. Equation (8) compares a sample x_i from condition A with a sample x_j from condition B and assigns $d_{ij}=1$ if $x_i > x_j$, $d_{ij}=-1$ if $x_i < x_j$, and $d_{ij}=0$ if $x_i=x_j$.

III. RESULTS

A. Directional localization

The directional localization results in terms of lateral angle are presented in Fig. 4(a) for the three stimulus conditions. The lateral angle data have been rounded to the nearest 5° and the data for all the subjects have been grouped together since all subjects showed similar responses in each condition. Each column shows the localization data grouped into the different distance ranges and each row contains the results for the different conditions. The size of the circle indicates the number of responses in a particular angular region, as described by the legend at the bottom of the figure. In all stimulus conditions, the localization data show that the lateral angles of the responses were close to the target lateral angles at the different distance ranges implying that there were no major lateral angle localization errors in all stimulus conditions. Figure 4(b) shows the mean absolute difference between the target and response lateral angle for the different distance ranges for the three stimulus conditions. In all conditions there is a statistically significant increase in absolute lateral angle difference at the 10–20 cm distance range but the effect is small (Table I). Comparing across the conditions, there is a statistically significant increase in the absolute lateral angle difference in the two DVF stimulus conditions compared to the intensity-scaled stimulus condition [$\chi^2(2,7600)=17.64$, $p<0.005$] but again the effect is small ($d=0.08$).



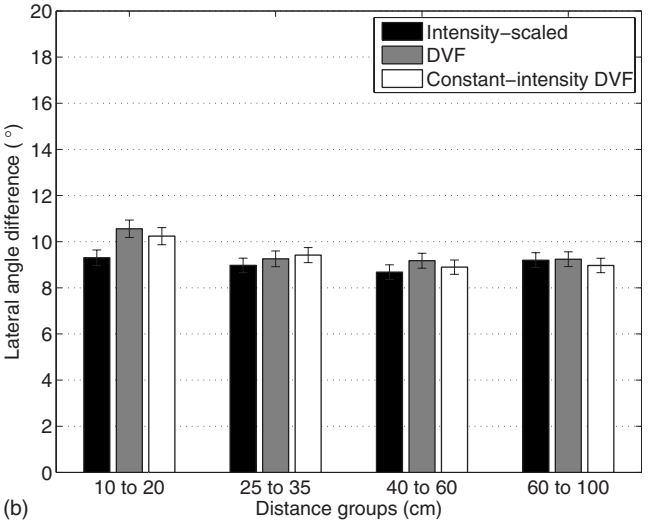
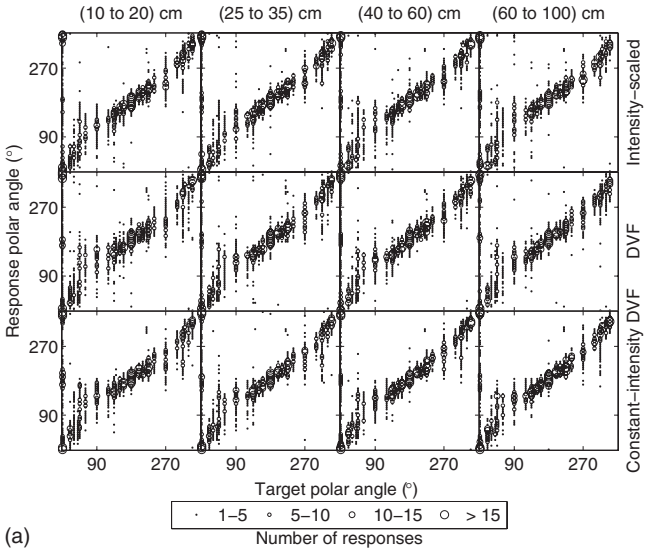


FIG. 4. (a) shows the directional localization results in terms of lateral angle. The data for the three stimulus conditions are shown in each row and are grouped into different distance ranges. (b) shows the mean absolute difference between the target and response lateral angle. The bars are grouped into different distance ranges and within each group, and the mean absolute angular difference is shown for the three stimulus conditions. The error bars indicate the 95% confidence interval of the mean.



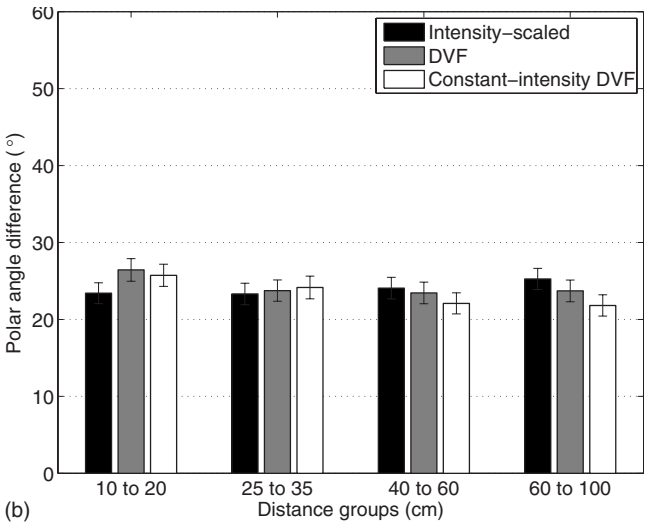


FIG. 5. (a) shows the directional localization results in terms of polar angle and (b) shows the mean absolute difference between the target and response polar angle. The data are shown in the same format as in Fig. 4.

The directional localization results in terms of polar angle are presented in Fig. 5(a), where the data are shown in a similar format as Fig. 4(a). The polar angles have been rounded to the nearest 10°. The data show that, in general, the polar angles of the responses were close to the target polar angles implying that there were no major polar angle localization errors in all stimulus conditions. Some of the localization differences between the target and response polar angles can be attributed to front-back confusions, where the lateral angle of the response is correct but the perceived direction is incorrectly placed in the wrong front-back hemisphere (Carlile *et al.*, 1997). In our analysis, a front-back confusion is identified when the response lateral angle is within 10° of the target lateral angle but the response is located in the wrong front-back hemisphere. The percentage of front-back errors for the different stimulus conditions are shown in Fig. 6. In all conditions, the percentage of front-back errors are of a similar order as free-field sound localization experiments using a head-pointing response paradigm reported elsewhere (Carlile *et al.*, 1997). The mean absolute difference between the target and response polar angle in each distance range for the three stimulus conditions is shown in Fig. 5(b), where front-back confusions have been removed. There are statistically significant increases in the

TABLE I. The results of a Kruskal–Wallis non-parametric ANOVA conducted on the absolute lateral angle differences across the distance ranges are shown for each stimulus condition. The table shows the χ^2, degrees of freedom (dof) and p-values. A *post hoc* analysis (Tukey HSD) revealed a statistically significant difference in localization performance between the 10–20 cm distance range and the other distance ranges in all stimulus conditions. The index measuring the size of this effect, d, is also shown in the table, and is small for all conditions.

Stimulus condition	χ^2	dof	p	d
Intensity-scaled	8.53	3	0.04	0.02
DVF	28.93	3	<0.005	0.08
Constant-intensity DVF	20.76	3	<0.005	0.06

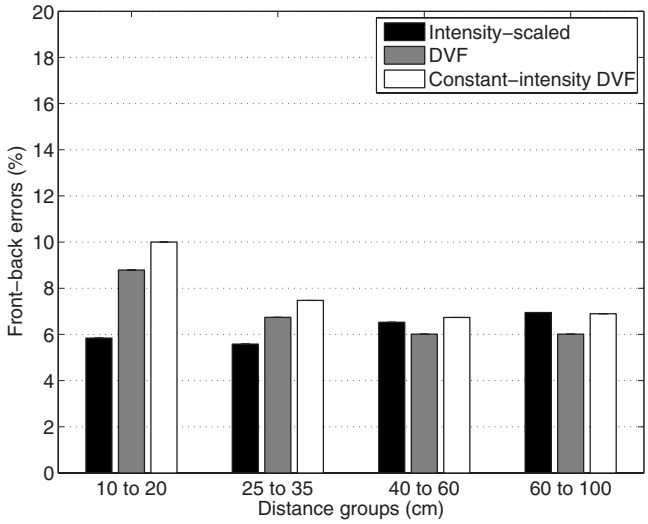


FIG. 6. The percentage front-back errors for the three stimulus conditions. The data are shown in the same format as in Fig. 4(b).

absolute polar angle difference between the target and response angle in the 10–20 cm distance range compared to the other distance ranges for the two DVF-generated stimulus conditions and in the 60–100 cm distance range for the constant-intensity stimulus condition though the effect is small (Table II). Comparing across the stimulus conditions, there are statistically significant differences in absolute polar angle localization performance at the different distance ranges but again the effect is small (Table III).

B. Distance perception

The perceived distances for the three stimulus conditions are shown in Fig. 7. Distances have been rounded to the nearest 5 cm. The data for both the left and right-hand sides have been grouped and divided into one of three azimuth regions shown in each column: front (0°–60°), side (60°–120°), and back (120°–180°), where 0° is directly in front. Each row shows the data for each stimulus condition. In all conditions, the localization data show a high degree of overestimation of the intended distance of the sound source indicated by the responses above the dashed line which indicates the theoretical responses for accurate distance perception. The mean perceived distance for all stimulus conditions separated into different distance ranges and azimuth regions

TABLE II. The results of a Kruskal–Wallis non-parametric ANOVA conducted on the absolute polar angle differences across the distance ranges are shown for each stimulus condition. The table shows the χ^2, dof and p-values. A *post hoc* analysis (Tukey HSD) revealed a statistically significant difference in localization performance in the 10–20 cm distance range compared to the other distance ranges for the two DVF-generated stimulus conditions and in the 60–100 cm distance range for the constant-intensity stimulus condition. The index measuring the size of these effects, d, is also shown in the table, and is small for all conditions.

Stimulus condition	χ^2	dof	p	d
Intensity-scaled	25.33	3	<0.005	0.02
DVF	27.28	3	<0.005	0.06
Constant-intensity DVF	58.01	3	<0.005	0.09

TABLE III. The results of a Kruskal–Wallis non-parametric one-way ANOVA conducted on the absolute polar angle differences across the stimulus conditions are shown for each distance range. The table shows the χ^2, dof and p-values. A *post hoc* analysis (Tukey HSD) revealed a statistically significant difference in localization performance between the constant-intensity stimulus condition and the DVF-generated stimulus conditions. The index measuring the size of these effects, d, is also shown in the table, and is small for all conditions.

Distance range	χ^2	dof	p	d
10–20 cm	7.84	2	0.02	0.05
25–35 cm	0.03	2	0.099	na
40–60 cm	11.74	2	<0.005	0.04
60–100 cm	49.53	2	<0.005	0.07

is shown in Fig. 8(a). In the intensity-scaled and DVF-generated stimulus conditions, the mean perceived distance follows the expected trend; that is, at the further distance ranges the mean perceived distance was larger than the closer distance ranges. However, the mean perceived distance is generally outside the intended distance range. In the constant-intensity DVF-generated stimulus condition, the mean perceived distance only fits the expected trend when the stimulus was located toward the side region. In the front and back regions, subjects generally perceived the stimulus as being at 50 and 65 cm, respectively, for all target distance ranges. From Fig. 7, it can also be seen that subjects exhibited a high amount of variability in perceived distance accuracy indicated by the spread of responses. The mean difference between the target and response distance for the different distance ranges and azimuth groups is shown in Fig. 8(b), where a positive difference indicates overestimation of the simulated distance. It can be seen that in all conditions subjects tended to overestimate the intended sound source distance when the stimulus was within 60 cm with slightly

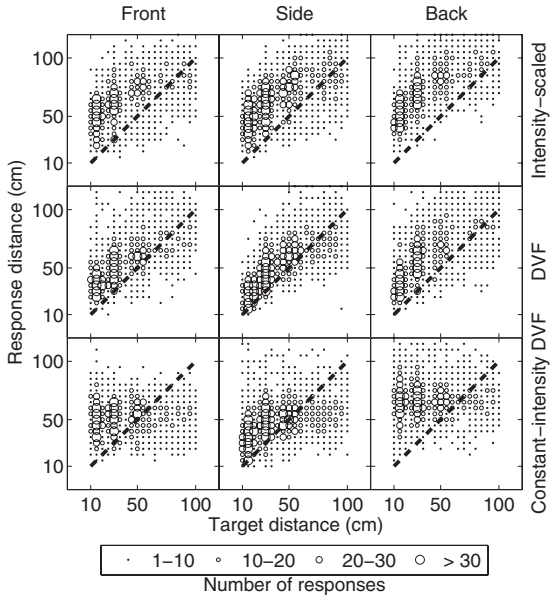
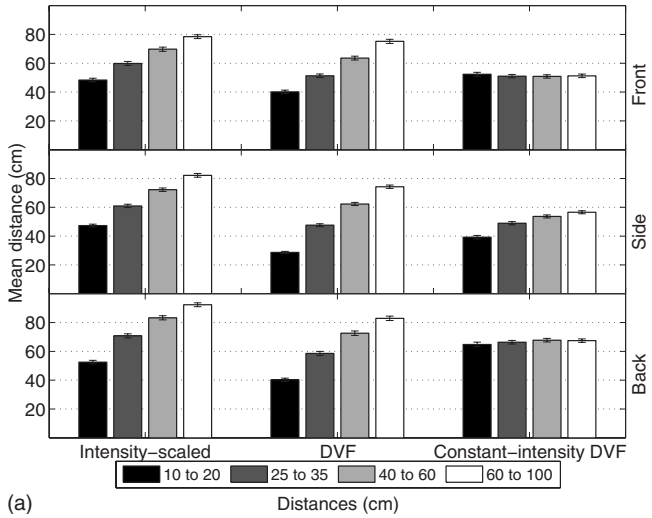


FIG. 7. The perceived distances are shown for the three stimulus conditions. Each column shows the data for one of three azimuth regions defined as front (0°–60°), side (60°–120°), and back (120°–180°), where 0° is directly in front. Each row shows the data for each stimulus condition. The dashed line indicates the theoretical responses for accurate distance perception.



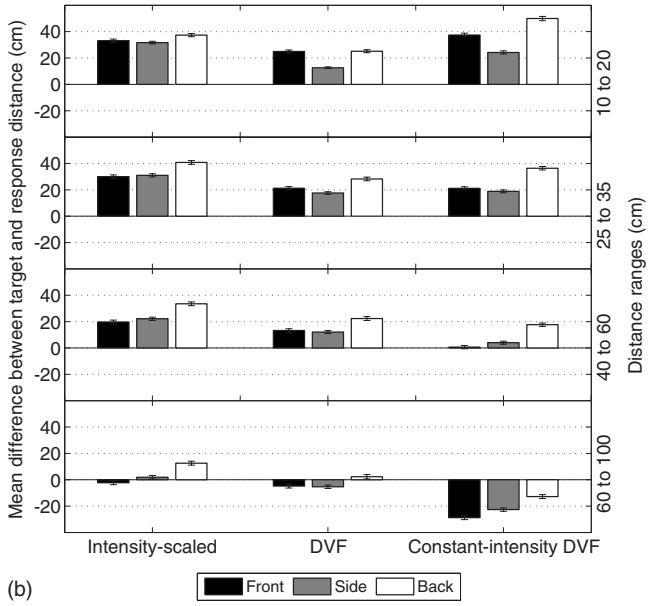


FIG. 8. (a) shows the mean perceived distance of each distance range for the three stimulus conditions. (b) shows the mean difference between the target and response distance, where the bars are grouped into the different azimuth regions and the results for the three stimulus conditions are shown within each group. The error bars indicate the 95% confidence interval of the mean.

greater overestimation occurring in the back region. The larger localization differences in the back region are probably due to the awkwardness of indicating perceived distances around the pole supporting the magnetic field transmitter of the sensor system, as well as the fact that the subject needs to turn his whole body first before indicating the perceived sound source location. Comparing the data for the front and side regions only, there is a statistically significant improvement in distance localization accuracy for sounds located to the side within 40 cm in the DVF-generated stimulus condition $[\chi^2(1,2650)=234.56,\ p<0.005]$, while there is no statistically significant improvement in the intensity-scaled stimulus condition in the same region $[\chi^2(1,2650)=0.84,\ p=0.36]$. In the constant-intensity DVF-generated stimulus condition, there is a statistically significant improvement in distance localization accuracy for sounds located to the side within the 10–20 cm distance range $[\chi^2(1,2650)=198.59,\ p<0.005]$. Comparing the DVF-generated stimulus condition and the intensity-scaled stimulus condition in Fig. 8(b), there is a noticeable improvement in distance perception for the DVF-generated stimulus condition for all azimuth regions when the sound source is located within 60 cm and calculation of the effect size index showed this improvement to be moderately large ($d=0.41$). The largest improvement in distance perception is in the side region within 40 cm of the subject ($d=0.61$).

IV. DISCUSSION

We conducted a psychoacoustic experiment to investigate the perceptual fidelity of near-field VAS generated using near-field HRTFs synthesized using the DVF method compared to that generated by simply scaling the intensity of the sound stimulus. A comparison of the distance perception from the DVF-generated stimulus condition and the intensity-scaled condition shows a statistically significant difference in localization performance in the side region between 60° and 120° and within 40 cm of the subject. In this region, subjects exhibited substantial overestimation of the sound source distance when intensity-scaled stimuli were used (on average, 35 cm from the intended sound source distance) compared to less than an average of 25 cm for the DVF-generated stimulus condition. This indicates that simply adjusting the intensity of the stimulus may not be enough to indicate to subjects that the sound stimulus is intended to be within 40 cm. Also, calculation of the effect size index showed that the improvement in distance perception using the DVF-generated stimulus over the intensity-scaled stimulus is large. The improvement in distance perception can be understood by studying the ILD introduced by the DVF. The ILD is calculated by taking the difference (in decibels) between the left and right DVFs for a particular sound source position. The magnitude of the ILD introduced by the DVF is shown as contour plots for different frequencies in Fig. 9. The DVF was calculated for a rigid sphere of 7 cm radius with "ears" located at ±100° azimuth from the mid-sagittal plane on the audio-visual horizon. Within each plot, the horizontal axis shows the distance between the sound source and the center of the head (on a logarithmic scale) and the vertical axis shows the lateral angle of the sound source. Areas of approximately equal ILD magnitude are indicated by color. A comparison of the ILD in the area of improved distance perception using the DVF-generated stimulus with that outside this region shows that the ILDs are greatest in the area between 60° and 120° within a distance of about 40 cm. Outside of this region, the ILDs are smaller. Hence, it would appear that the improvement in distance localization performance is related to the ILD introduced by the DVF. This ILD should approximate the true ILD of the near-field HRTF since the size of the rigid sphere used to calculate the DVF is similar in size to the subject's head.

In the case where intensity was removed as a cue for distance in the constant-intensity DVF-generated stimulus condition, it was observed that there was some distance discrimination in the side region and within the 10–20 cm distance range, though performance was poor. It would appear

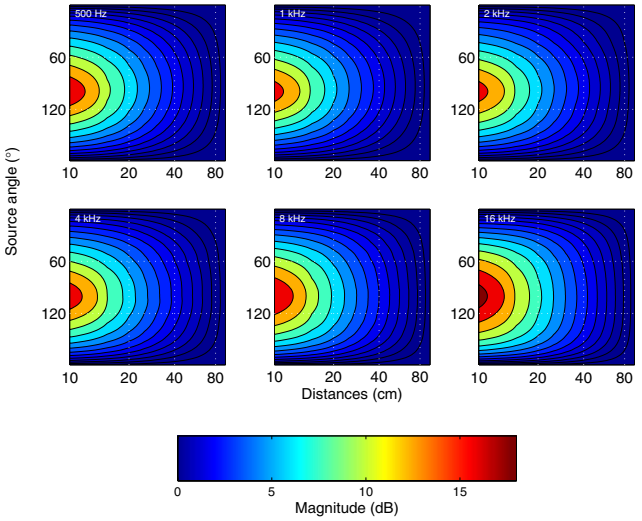


FIG. 9. (Color online) The ILD introduced by the DVF as a function of sound source distance and lateral angle is shown as contour plots for different frequencies. Distances are shown on a logarithmic scale. The ILD is calculated as the difference (in dB) between the left and right DVFs for a particular sound source position. The DVF was calculated for a rigid sphere of 7 cm radius with ears located at ±100° azimuth from the mid-sagittal plane on the audio-visual horizon. Color is used to indicate areas of approximately equal magnitude in the ILD.

that the ILD introduced by the DVF provides some cue for absolute distance discrimination but is not strongly salient. Hence, in using the DVF method to synthesize near-field HRTFs for generating VAS, the combination of intensity cues and ILD cues seem to be important for better distance perception.

Although there is an improvement in distance perception in the DVF-generated stimulus condition compared to the intensity-scaled stimulus condition, there is still a high degree of overestimation of the distance and variability for all subjects. However, a recent summary of human distance perception research by Zahorik *et al.* (2005) shows that it is quite common for listener's to underestimate distance of far-field sources and overestimate the distance of near-field sounds. They also show that there is often high variability in judgments of sound source distance. Hence, it is difficult to determine whether the overestimation and variability in distance perception are due to the generated VAS or due to the generally poor distance discrimination by humans.

It is also important to ensure that directional localization accuracy is not degraded by the synthesized near-field stimulus. In all conditions, there was a statistically significant difference in the mean difference between the target and response lateral and polar angles for different distance ranges. In particular, there was a statistically significant increase in the lateral and polar angle differences at the 10–20 cm distance range. However, calculation of the effect size index showed the effect of the differences in directional localization performance between the different distance ranges to be small for all conditions. It can be expected that directional localization errors would increase with decreasing sound source distance since a small error in the placement of the electromagnetic sensor to indicate the perceived location will translate into a larger angular difference at a close distance compared to a farther distance.

The advantage of the DVF technique is that it avoids the technical difficulties associated with measuring near-field HRIRs on human subjects which has hindered the development of high fidelity near-field VAS. There have been some other methods presented in the literature for generating near-field VAS without having to measure near-field HRIRs on human subjects. Martens and Yoshida (2000) presented a method for synthesizing near-field VAS using an approximation of the rigid sphere model. In their work a number of signal processing steps are applied to the stimulus to synthesize a sound source in the near-field. First, the ITD for the contralateral ear signal is adjusted for the intended near-field distance. Second, attenuation due to distance is applied to both the signals for both ears equally and an additional attenuation is applied to the contralateral ear signal to mimic the ILD in the near-field. Finally, a single-pole, single-zero filter that approximated the response of a rigid sphere is applied to model the frequency-dependent acoustical response of the head. Their method is based on the results of a simple psychoacoustic experiment that showed that the perceived distance of speech stimuli could be changed by manipulation of the ILD. However, the accuracy of the distance localization of the stimuli was not directly examined. It is also unclear whether directional localization accuracy would be affected by such adjustment of the ILD. The advantage of the DVF method over the method proposed by Martens and Yoshida (2000) is that the DVF is able to account for the ILD for sounds in the near-field, as well as low-frequency parallax effects of the HRTF. Since a human head and a rigid sphere of the same size have a similar low-frequency response, the DVF calculated from the rigid sphere model describes the change in the low frequencies of the HRTF as the sound source distance is changed from the far-field to the near-field. However, the DVF is unable to account for high-frequency parallax effects since pinnae are not modeled by the rigid sphere model. Although Møller *et al.* (1995) and Kim *et al.* (2001) suggested that it is possible to simulate a source in the near-field by choosing the appropriate far-field HRTF according to a simple parallax model, it is unclear that the high-frequency spectra of the HRTFs indeed do change in a manner that follows the simple parallax model they have presented and whether directional localization accuracy will be affected.

Brungart and Simpson (2001) conducted a psychoacoustic experiment to evaluate human sound localization performance of near-field VAS generated using near-field HRIRs measured on an acoustic manikin. They found that, in general, listener's could discriminate sound source distance but directional localization performance was not as accurate as that of nearby sounds in the free-field. This is not surprising since Wenzel *et al.* (1993) had previously shown that virtual sound stimuli generated using HRIRs recorded on a subject's own ears caused less directional localization errors than that generated using HRIRs measured on another person or a manikin. Brungart and Simpson (2001) investigation highlights the need for HRIRs measured on the subject for the generation of high fidelity near-field VAS. In our work, HR-

IRs measured on the individual subjects have been used to synthesize the near-field HRIRs and directional localization is preserved.

Duraiswami *et al.* (2004) presented a method for synthesizing near-field HRIRs from HRIRs measured in the far-field. By representing each measured HRTF as a series of multipole solutions to the Helmholtz wave equation and solving for the coefficients of the multipole solutions simultaneously, the HRTFs for any direction and distance from the head can theoretically be evaluated from the obtained coefficients. The major drawback of this technique is that close to 2000 far-field HRIRs need to be measured to accurately evaluate HRTFs up to 16 kHz. In our work, we have synthesized near-field HRIRs from far-field HRIRs measured on the subject using a rigid sphere model with the same radius as the head of the subject. Our method does not increase the number of far-field HRIRs that need to be measured to generate the near-field VAS.

V. CONCLUSION

We have presented a method for synthesizing near-field HRTFs using a rigid sphere model of the human head by calculating a DVF. The DVF approximates the ILD cues of the HRTF as a function of sound source distance. Using this method, the technical difficulties associated with measuring near-field HRTFs on human subjects are avoided. A human sound localization experiment was conducted to test the perceptual fidelity of the near-field VAS generated by this method as well as VAS generated by simply adjusting the intensity of the stimulus. Results show improved distance perception for sounds at simulated distances of up to 60 cm using the DVF compared to simple intensity adjustment, while maintaining directional accuracy, with the largest improvements for sound sources located to the side and within 40 cm. When intensity cues are removed from the DVF-generated stimulus, results still showed some discrimination of sound source distance but distance perception accuracy was poor.

ACKNOWLEDGMENTS

The DVF method is patent pending and thanks is given to Personal Audio Pty. Ltd. for permission to publish these results.

Blauert, J. (**1997**). *Spatial Hearing* (MIT, Cambridge, MA).
Brungart, D. (**1999**). "Auditory localization of nearby sources. III. Stimulus effects," J. Acoust. Soc. Am. **106**, 3589–3602.
Brungart, D. (**2002**). "Near-field virtual audio displays," Presence: Teleoperators & Virtual Environments **11**, 93–106.
Brungart, D., and Rabinowitz, W. (**1999**). "Auditory localization of nearby sources. Head-related transfer functions," J. Acoust. Soc. Am. **106**, 1465–1479.
Brungart, D. S., and Simpson, B. D. (**2001**). "Auditory localization of nearby sources in a virtual audio display," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, pp. 107–110.
Brungart, D., Durlach, N., and Rabinowitz, W. (**1999**). "Auditory localization of nearby sources. II. Localization of a broadband source," J. Acoust. Soc. Am. **106**, 1956–1968.
Brungart, D., Rabinowitz, W., and Durlach, N. (**2000**). "Evaluation of response methods for the localization of nearby objects," Percept. Psychophys. **62**, 48–65.
Carlile, S. (**1996**). *Virtual Auditory Space: Generation and Application* (R. G. Landes Company, Austin, TX).
Carlile, S., Leong, P., and Hyams, S. (**1997**). "The nature and distribution of errors in the localization of sounds by humans," Hear. Res. **114**, 179–196.
Carlile, S., Jin, C., and van Raad, V. (**2000**). "Continuous virtual auditory space using HRTF interpolation: Acoustic and psychophysical errors," in First IEEE Pacific-Rim Conference on Multimedia, International Symposium on Multimedia Information Processing, Sydney, Australia.
Cliff, N. (**1993**). "Dominance statistics: Ordinal analyses to answer ordinal questions," Psychol. Bull. **114**, 494–509.
Duda, R. O., and Martens, W. L. (**1998**). "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**, 3048–3058.
Duraiswami, R., Zotkin, D. N., and Gumerov, N. A. (**2004**). "Interpolation and range extrapolation of HRTFS," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP, Montreal, QC Canada, Vol. **4**, pp. 45–58.
Hartung, K., Braasch, J., and Steinbing, S. J. (**1999**). "Comparison of different methods for the interpolation of head-related transfer functions," in AES 16th International Conference, Rovaniemi, Finland, pp. 319–328.
Hammershøi, D., and Møller, H. (**1996**). "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427.
Kim, H., Suzuki, Y., Takane, S., and Sone, T. (**2001**). "Control of auditory distance perception based on the auditory parallax model," Appl. Acoust. **62**, 245–270.
Kuhn, G. F. (**1977**). "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157–167.
Martens, W., and Yoshida, A. (**2000**). "Psychoacoustically-based control of auditory range: Display of virtual sound sources in the listener's personal space," in Proceedings of the International Conference on Information Society in the 21st Century: Emerging Technologies and New Challenges (IS2000) Aizu-Wakamatsu, Japan.
Møller, H. (**1992**). "Fundamentals of binaural technology," Appl. Acoust. **36**, 171–218.
Møller, H., Hammershøi, D., and Jensen, C. (**1995**). "Binaural synthesis, head-related transfer functions, and uses thereof," Patent No. WO 95/23493.
Rabinowitz, W. M., Maxwell, J., Shao, Y., and Wei, M. (**1993**). "Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere," Presence: Teleoperators & Virtual Environments **2**, 125–129.
Wenzel, E., and Foster, S. (**1993**). "Perceptual consequences of interpolating head-related transfer functions during spatial synthesis," in Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 102–105.
Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.
Zahorik, P., Brungart, D., and Bronkhorst, A. (**2005**). "Auditory distance perception in humans: A summary of past and present research," Acta. Acust. Acust. **91**, 409–420.
Zhou, B., Green, D., and Middlebrooks, J. (**1992**). "Characterization of external ear impulse responses using Golay codes," J. Acoust. Soc. Am. **92**, 1169–1171.

Wind noise in hearing aids with directional and omnidirectional microphones: Polar characteristics of behind-the-ear hearing aids

King Chung[a)]
Department of Allied Health and Communicative Disorders, Northern Illinois University, DeKalb, Illinois 60115

Luc Mongeau[b)]
Department of Mechanical Engineering, McGill University, Montreal, Quebec H3A 2K6, Canada

Nicholas McKibben[c)]
Department of Mechanical Engineering, Purdue University, West Lafayette, Indiana 47907

(Received 29 September 2008; revised 27 January 2009; accepted 28 January 2009)

Wind noise can be a significant problem for hearing instrument users. This study examined the polar characteristics of flow noise at outputs of two behind-the-ear digital hearing aids, and a microphone mounted on the surface of a cylinder at flow velocities ranging from a gentle breeze (4.5 m/s) to a strong gale (22.5 m/s) . The hearing aids were programed in an anechoic chamber, and tested in a quiet wind tunnel for flow noise recordings. Flow noise levels were estimated by normalizing the overall gain of the hearing aids to 0 dB. The results indicated that the two hearing aids had similar flow noise characteristics: The noise level was generally the lowest when the microphone faced upstream, higher when the microphone faced downstream, and the highest for frontal and rearward incidence angles. Directional microphones often generated higher *flow* noise level than omnidirectional microphones but they could reduce far-field background noise, resulting in a lower *ambient* noise level than omnidirectional microphones. Data for the academic microphone-on-cylinder configuration suggested that both turbulence and flow impingement might have contributed to the generation of flow noise in the hearing aids. Clinical and engineering design applications are discussed. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3086268]

PACS number(s): 43.66.Ts, 43.28.Vd, 43.60.Fg, 43.38.Kb [BLM] Pages: 2243–2259

I. INTRODUCTION

Wind noise can be an annoyance for hearing instrument users, especially for those who enjoy outdoor activities. Excessive wind noise often results in complaints of poor sound quality and/or inability to communicate. A hearing instrument's ability to manage wind noise is reported to affect the satisfaction and the frequency of instrument usage (Kochkin, 2000, 2005).

Although wind noise has been a significant problem, there are only few reports on the characteristics and the management of wind noise in literature. Grenner *et al.* (2000) measured the level and spectra of wind noise at the output of four behind-the-ear hearing aids at 0° at a velocity of 7.0 m/s. They reported that wind noise had the highest amplitude at low frequencies, and that its characteristics differed among hearing aids. Thompson and Dillon (2002) measured the wind noise at the outputs of directional and omnidirectional microphones mounted on different styles of hearing aids and worn in the ears of a Knowles Electronic Mannequin for Acoustic Research (KEMAR) (Burkhard and Sachs, 1975). They concluded that directional microphones are, in general, noisier than omnidirectional microphones, and that wind noise is relatively lower at head angles of +90° and −90°.

Beard and Nepomuceno (2001) measured wind noise of an omnidirectional microphone and two directional microphones mounted on the faceplate of an in-the-ear hearing aid worn on a KEMAR in a wind tunnel. They also reported that directional microphones yielded higher wind noise levels. These studies, however, were limited to relatively low velocities (i.e., 5 m/s in Thompson and Dillon, 2002; and 2.1 and 5 m/s in Beard and Nepomuceno, 2001). Furthermore, no detailed analysis of the polar, spectral, or temporal characteristics of wind noise was reported.

The characteristics of wind noise in hearing aids can be affected by many factors such as venting, case/shell design, microphone location, signal processing algorithms, frequency response of the hearing aid, and wind velocity. In order to examine the cumulative effects of these factors on the characteristics of wind noise in users' ear canals, wind noise was recorded at the outputs of two commercially available behind-the-ear digital hearing aids. The goals were to investigate the effects of head angle and microphone type at three wind velocities [i.e., from a gentle (4.5 m/s) to a strong breeze (13.5 m/s)] and to examine the similarities and differences between the two behind-the-ear hearing aids. Measurements were also made on one of the hearing aids at higher wind velocities, ranging from a fresh gale (18.0 m/s) to a

a)Electronic mail: kchung@niu.edu
b)Electronic mail: luc.mongeau@mcgill.ca
c)Electronic mail: nick.mckibben@gmail.com

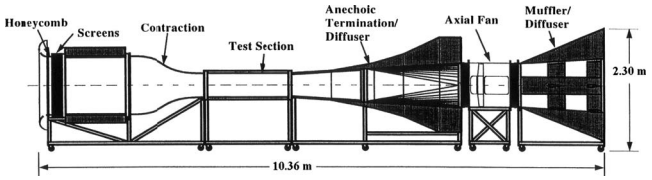


FIG. 1. Schematic of the wind tunnel.

strong gale (22.5 m/s). Additionally, as the aerodynamic pattern of a flow passing a cylindrical body is a well studied phenomenon, the wind noise recorded by a microphone mounted on the surface of a rigid cylinder was investigated to shed light on factors contributing to the generation of wind noise in hearing aids.

Note that "wind" normally refers to atmospheric air motion. The term "flow" is used hereafter to refer to the air flow generated in the wind tunnel. This is the first of a series of studies to examine the polar, spectral, and temporal characteristics of wind noise in hearing aids with directional and omnidirectional microphones.

II. METHODS

Each digital hearing aid was programed in an anechoic chamber when worn on KEMAR's right ear. The hearing aid and the KEMAR head were then transported to an acoustically treated wind tunnel located in a nearby facility. The noise at the hearing aid output was recorded for head angles from 0° to 360° in 10° increments. Measurements were made in the absence of a flow and at three flow velocities with the hearing aid programed to directional (DIR) and omnidirectional (OMNI) microphone modes. The KEMAR head was then transported back to the anechoic chamber and the hearing aid output in response to a 75 dB SPL pink noise was recorded with identical equipment and volume control settings as the wind tunnel measurements. This recorded noise was later used as the reference to estimate flow noise levels.

A. Wind tunnel characteristics

1. Construction

Figure 1 shows a simplified schematic of the wind tunnel. The Eiffel type wind tunnel had a closed, rectangular test section. A variable speed axial flow fan located downstream of the test section generated the flow. The Plexiglas test section walls were approximately 52.7 cm wide × 60.3 cm high × 121.9 cm long. A wire mesh screen was installed at the end of the test section to protect the fan against any debris.

Several wind tunnel features were designed to ensure a uniform flow in the test section: (1) a honeycomb flow straightener at the wind tunnel inlet, (2) several layers of wire mesh to break down large flow structures into fine-grained flow, and (3) a smooth contraction section to gradually accelerate the flow into the test section, ensuring a uniform velocity profile with thin boundary layers. The flow velocity uniformity at the entrance of the test section was within 1%, with velocity fluctuations less than approximately 3%. Two progressive diffuser sections were located downstream of the fan to minimize flow resistance and ensure a stable flow (Brown and Mongeau, 1995).

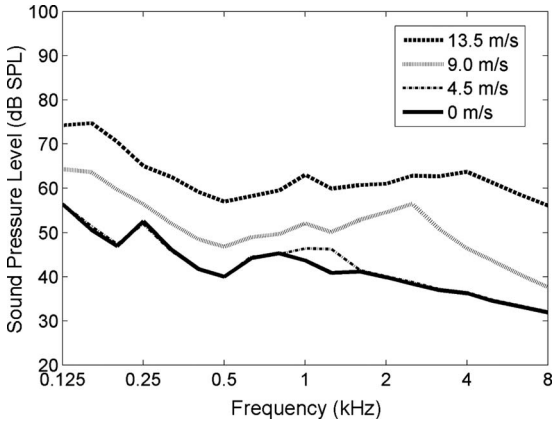


FIG. 2. One-third octave band ambient noise in the wind tunnel with no flow (0 m/s) and at three flow velocities. The flow noise was measured using a microphone with 1.27 cm diameter covered by an elliptical windscreen.

The wind tunnel was also treated to reduce acoustic noise in the test section. The inlet and diffuser sections were lined with sound absorbing materials in order to minimize acoustic noise from the fan and other aerodynamic noise sources. An additional muffler/diffuser was also implemented downstream of the axial fan section to reduce motor/fan noise and to reduce the noise in the surrounding area entering the wind tunnel. The ambient noise floor of the wind tunnel was measured using a microphone with 1.27 cm diameter (ER-11 in. microphone, Etymotic Research), which was protected with an ellipsoidal windscreen (B&K UA 0781). The ambient tunnel noise levels at 0, 4.5, 9.0, and 13.5 m/s were measured when the microphone was located approximately 30 cm upstream of the KEMAR head. The spectra of the ambient noise analyzed in one-third octave bands from 125 to 8000 Hz are shown in Fig. 2. Other details of the wind tunnel construction and flow characteristics are available in Brown and Mongeau (1995).

2. Flow velocity calibration

The flow velocity was altered by changing the fan rotational speed, and measured by using a Pitot tube. The Pitot tube was placed approximately 30 cm upstream from the location of the KEMAR head in the test section. The Pitot tube data were collected at fan speeds from 8.5 to 24.5 Hz in 4 Hz increments. Since fan speed and flow velocity data were linearly related, the desired fan speeds for target flow velocities were calculated using a linear regression equation. The fan speeds were determined to be 10.3, 19.3, and 28.2 Hz in order to obtain flow velocities of 4.5, 9.0, and 13.5 m/s, respectively. These velocities are equivalent to 10.0, 20.1, and 30.1 miles/hour, respectively. The Pitot tube was removed during flow noise measurements to reduce vortex shedding noise. These flow velocities yielded ratings of 3 (gentle breeze), 5 (fresh breeze), and 6 (strong breeze) on the 13-point Beaufort scale, respectively. The lowest flow velocity was estimated to be the average wind velocity on a normal non-windy day in the United States (National Climate

Data Center, 2006a). The highest wind velocity is the threshold for the National Weather Service to issue a wind advisory (National Climate Data Center, 2006b).

B. Hearing aid characteristics

Two behind-the-ear digital hearing aids were used, designated as BTE1 and BTE2. BTE1 had 14 signal processing channels and BTE2 had 16 channels, and they were made by two different manufacturers. Both hearing aids were worn on KEMAR's large pinnae on the right side and their outputs were coupled into KEMAR's ear canal via a No. 13 tubing (inner diameter of 1.9 mm) and an unvented skeleton earmold. Neither hearing aid had wind noise detection or a signal processing algorithm to manage wind noise.

1. Microphones

The directional microphones of both hearing aids had non-adaptive polar patterns and they utilized a dual microphone design in which the outputs of two omnidirectional microphones were processed to form one directional microphone. Both manufacturers reported that the directional microphones had hypercardioid patterns when measured in free field (Chung, 2004). When programed to the OMNI mode, the front microphone was active and the back microphone was disabled for both hearing aids.

2. Signal processing characteristics

Typical digital hearing aids often have several algorithms/limiters that can affect the input-output functions of the hearing aids: (1) a broadband input limiter to limit the overall level of the input; (2) an expansion algorithm to reduce low level noise, typically microphone noise and/or circuit noise below 35 dB SPL; (3) a wide dynamic range compression (WDRC) algorithm that has a linear region at low input level (e.g., $<$45 dB SPL), a compression region at low to high input levels (e.g., 45–90 dB SPL), and an output limiting region at high input levels (e.g., $>$90 dB SPL); and (4) a broadband output limiting algorithm to limit the overall level of the hearing aid output. The broadband input limiter could be an algorithm to limit the input signal or simply the upper limit imposed by the digital chip of the hearing aid. In addition, the expansion and WDRC algorithms are typically implemented in each frequency channel (i.e., narrowband), whereas the broadband input and output limiters usually act on the overall/summed signal.

In this study, both hearing aids had the expansion and WDRC algorithms in multiple channels. The compression region of the WDRC algorithms was linear. BTE1 had a broadband automatic gain control *output* limiting algorithm, which reduced the gain of the overall signal by changing the instantaneous gain (e.g., if the overall level would exceed the predetermined level by 5 dB at any time, the algorithm would reduce the gain of the overall signal by 5 dB). BTE2 had a broadband *input* compression limiting algorithm with a compression ratio of 8:1 (i.e., a change of 8 dB in the incoming signal resulted in a change of 1 dB at the output). All other signal processing features, such as feedback suppression and noise reduction algorithms, were disabled in both hearing aids.

3. Frequency response programming

The hearing aids were programed to be linear with relatively flat frequency responses when worn on KEMAR (i.e., relatively flat real-ear aided frequency responses). Pink noise was played from computer 1 and then fed to a powered loudspeaker (Mackie HR824) in an anechoic chamber. The overall sound pressure level of the pink noise was 70 dB SPL measured 1 m from the loudspeaker using a Type I sound level meter (Quest Precision 1200). The Mackie loudspeaker had a flat frequency response from 39 to 20 000 Hz ($\pm$1.5 dB) and a maximum output of 100 dB SPL. The KEMAR head was placed on a tripod with the ears at the same height as the center of the loudspeaker diaphragm.

The hearing aid output was recorded using a microphone with 1.27 cm diameter (ER-11) placed in the medial opening of a Zwislocki coupler. The signal was then fed to an external sound card (Sound Blaster Extigy) and analyzed simultaneously using SPECTRAPLUS software (Pioneer Hill Software) in another computer (computer 2, Dell D600). One-third octave band levels for pink noise (i.e., equal levels from 1 to 8000 Hz) were displayed on the computer screen. The frequency response of the hearing aid was adjusted to match that of the pink noise as much as possible. Hearing aid adjustments were accomplished using the NOAH programming software (Himsa) via a HIPRO programming interface and computer 2. The NOAH software and HIPRO interface are platforms that can program hearing aids from multiple manufacturers.

For each hearing aid, the OMNI frequency response was programed first. The frequency response of DIR was then adjusted to be as close to that of OMNI as possible. The low-frequency region of DIR was compensated to offset the inherent first-order high-pass-filter effects of directional microphones. If a match in frequency response was not possible, the OMNI frequency response was adjusted again so that the frequency response disparity between OMNI and DIR of the same hearing aid or across the hearing aids was minimal. This match was intended to minimize the effects of frequency response on flow noise measurements so that the levels obtained could be compared without additional signal processing or level adjustment.

The frequency responses of BTE1 OMNI and DIR modes were matched within 2.2 dB in one-third octave bands from 100 to 8000 Hz and that of BTE2 were matched within 3.7 dB (Fig. 3). Notice that BTE2 OMNI had a relatively higher level than BTE2 DIR below 250 Hz and relatively lower level between 250 and 1000 Hz. This pattern had also been noticed in other hearing aids and the disparity could not be eliminated (i.e., reducing the gain of OMNI at the lowest frequency would reduce the difference below 250 Hz but increase the difference between 250 and 1000 Hz). Attempts to adjust the frequency response of BTE2 DIR produced similar results. The frequency responses shown in Fig. 3 were obtained from attempts to minimize the differences at 125 and 500 Hz. The disparity was treated as intrinsic differences between the two microphone modes in real-life hear-

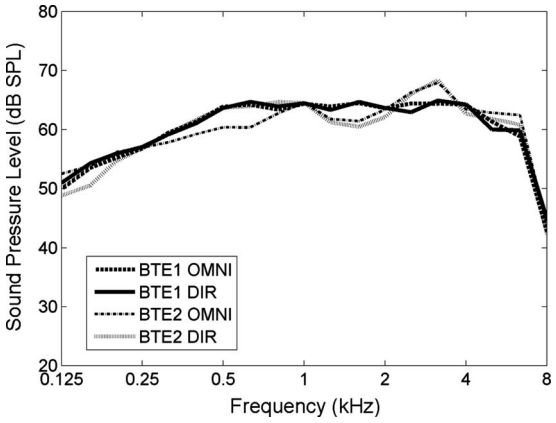


FIG. 3. Frequency responses of BTE1 and BTE2 in response to a 75 dB SPL pink noise presented at 0° azimuth in the anechoic chamber.

ing aid fitting and, therefore, no level adjustment was made in the flow noise level calculations. Overall, the differences between BTE1 and BTE2 for both OMNI and DIR were 3.7, 2.3, and 1.5 dB at 125, 500, and 2000 Hz one-third octave bands, respectively.

C. Flow noise measurements

After a hearing aid was programed in the anechoic chamber, the KEMAR head was transported to the wind tunnel for flow noise measurements. During transportation, the hearing aid was kept in KEMAR's ear and physical contact with other objects was avoided. This procedure ensured that the frequency responses used in the flow noise recordings were identical to those tested in the anechoic chamber.

1. BTE1 and BTE2

In the wind tunnel, the KEMAR head was mounted on a wooden rotating platform in the test section through a series of PVC pipe connections [Fig. 4(A)]. KEMAR's ears were placed at the center of the test section with respect to length, width, and height. The center reference notch on KEMAR's neck was set right above the 0° mark on the rotating platform, which was marked from 0° to 350° in 10° increments. A point equidistant to the side walls of the test section served as the reference for turning the KEMAR head. The angles reported in this study are variable head angles relative to a fixed flow direction.

The recording equipment used in flow noise measurements was identical to that used to program the hearing aid. To ensure optimal dynamic range, the maximum hearing aid overall output level was checked prior to flow noise recording. With the flow velocity at 13.5 m/s and the hearing aid set to DIR, the hearing aid output was monitored using ADOBE AUDITION 1.0 while the KEMAR head was rotated slowly from 0° to 360°. When the location corresponding to the highest flow noise output was identified, the recording level of the sound card was adjusted so that the hearing aid output peaked within 15 dB below the maximum amplitude level in AUDITION. This recording level was noted and maintained for all flow noise measurements.

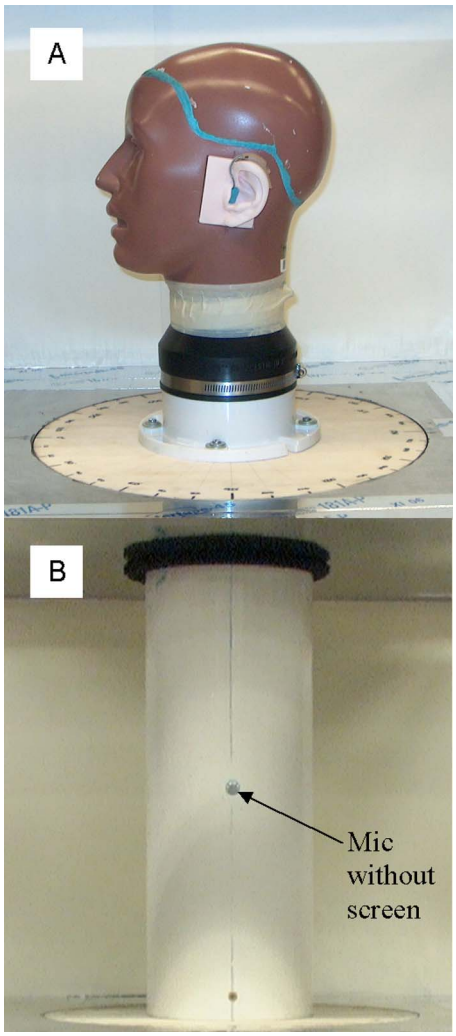


FIG. 4. (Color online) (A) KEMAR head on the rotating platform, and (B) a microphone mounted on a cylinder in the wind tunnel test section. (Note: Flow went from the left to right. A hearing aid worn on KEMAR's left ear and a microphone rotated to 270° are shown.)

Thirty-second recordings were made in 10° increments at four flow velocities, with the 0° angle repeated and recorded as 360° (i.e., 37 measurements per flow velocity). The hearing aid was later programed to OMNI and recordings were repeated for each flow velocity.

2. Repeatability and variability

Repeatability was verified by comparing the levels of consecutive recordings made using BTE2 DIR at each flow velocity, with identical equipment and hearing aid setup.

Variability was verified through repeated measurements using BTE1 OMNI at flow velocities of 4.5 and 13.5 m/s and BTE1 DIR at 0 and 9.0 m/s 14 days after the initial measurements. The hearing aid and earmold were taken off KEMAR's pinna and all the recording equipment was disassembled between the two measurements.

3. Left-right comparison

The flow noise at the outputs of BTE1 in DIR and OMNI modes was also recorded when BTE1 was worn on

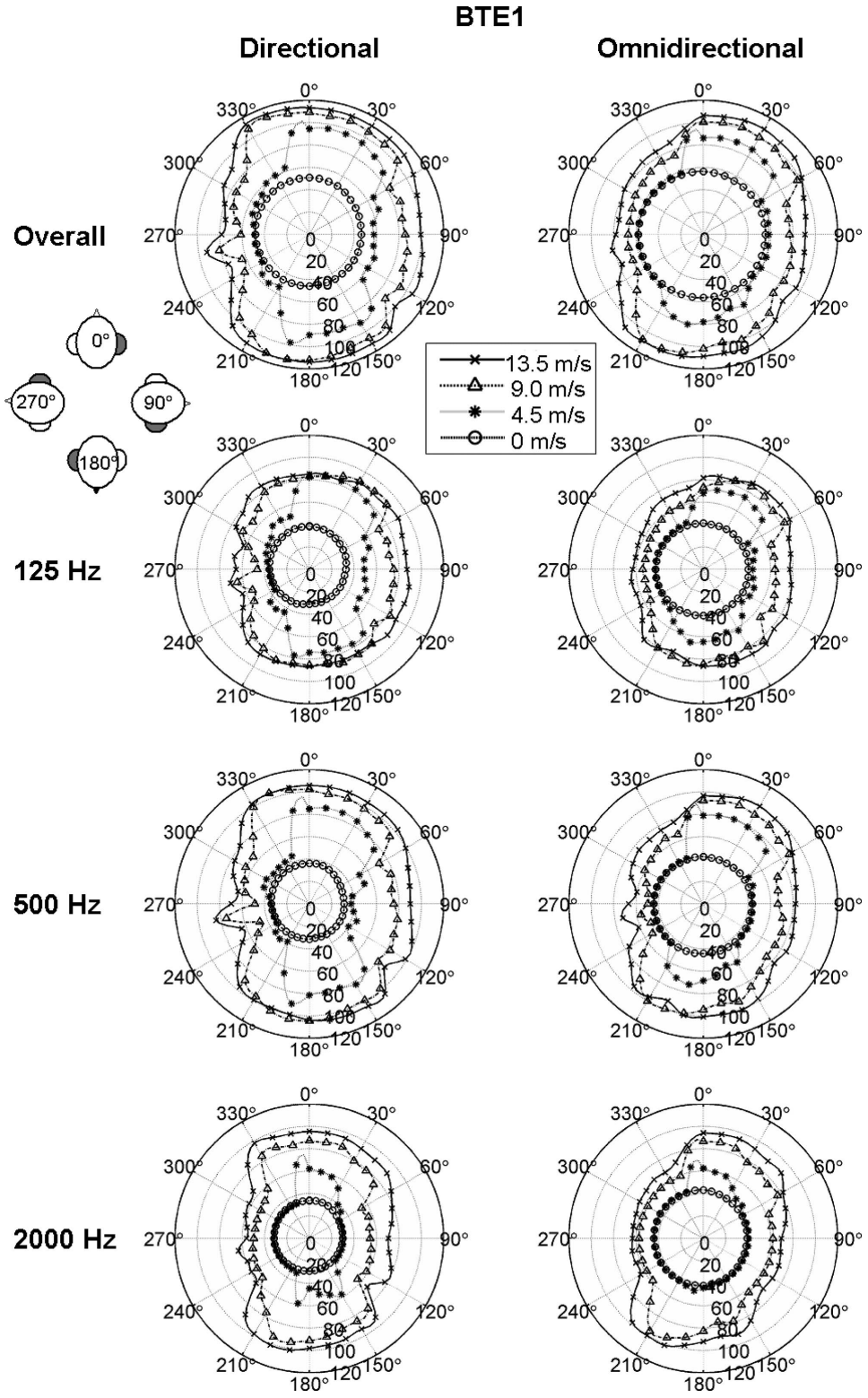


FIG. 5. Overall and one-third octave band levels of flow noise measured at the output of BTE1, which was set to the directional (DIR) and omnidirectional (OMNI) modes. The four aerial views of the KEMAR head with a darkened ear indicated the hearing aid locations when the head was turned at 0°, 90°, 180°, and 270°.

KEMAR's left ear (BTE1_Left). The frequency responses of BTE1_Left in DIR and OMNI modes matched those of BTE1 worn on KEMAR's right side within 1.9 dB from 100 to 6300 Hz, with most of the differences within 1 dB. The flow noise at the output of BTE1_Left was then compared with that of BTE1 in Fig. 5 in mirror images (i.e., flow noise at 60° and 90° recorded from BTE1 worn on the right was compared with the flow noise at 300° and 270° recorded from BTE1_Left, respectively).

Recordings were also made at 18.0 and 22.5 m/s (i.e., 40

and 50 miles/h, respectively) to examine the effects of the limiters in the hearing aid. The corresponding fan speeds were 37.2 and 46.1 Hz.

4. Flow around cylinder

The output of a 1.27 cm microphone (B&K 4936) mounted on a 15-cm diameter PVC cylinder was recorded at 0, 2.3, 4.5, 9.0, 13.5, and 18.5 m/s [Fig. 4(B)]. When the cylinder was set at 0° azimuth, the microphone was directly above the 90° mark on the platform (i.e., the microphone was on the right side of the cylinder). The microphone's protective grid was removed prior to mounting. The diaphragm of the microphone and the tack seal were flush with the surface of the cylinder. The microphone was located at the midspan of the cylinder, which extended over nearly the full height of the test section. Foam [black in Fig. 4(B)] was used to connect the top of the cylinder to the ceiling of the test section in order to ensure a good seal between the cylinder and the test section upper wall.

The microphone output was fed to a Quest Precision 1200 type I sound level meter via an angle-changing cable adaptor (B&K UA0122). Then the electrical output of the sound level meter was recorded using AUDITION. Other recording procedures were identical to those described above for hearing aids. The level range of the sound level meter was set to 70–140 dB SPL for recording flow noise at velocities of 4.5, 9.0, 13.5, and 18.5 m/s and to 30–100 dB SPL at flow velocities of 0 and 2.3 m/s.

D. Calibration of sound pressure levels

Immediately after each set of recordings was made, the KEMAR head and the recording equipment were transferred back to the anechoic chamber and a calibration noise was recorded.

1. Hearing aids

For each hearing aid, the output in response to a 75 dB SPL pink noise was recorded using identical hearing aid and recording equipment settings to those used in the wind tunnel recordings. This recorded noise served as the broadband 75 dB SPL calibration reference for the particular hearing aid. Adjustment of the hearing aid output in response to a 75 dB SPL pink noise input to 75 dB SPL essentially normalized the overall gain of the hearing aid to 0 dB, i.e., hearing aid output (assumed 75 dB SPL)=hearing aid input(75 dB SPL)+overall gain (0 dB).

This calibration procedure allowed the estimation of flow noise at the hearing aid input, and comparisons between flow noise levels among different hearing aids, but the estimated level depends on the frequency response of the hearing aid and on the gain at the particular frequency region. This disadvantage, however, was overcome by matching the frequency response described in the Hearing Aid Fitting section.

2. Flow around cylinder

The microphone was dismounted from the cylinder after making the flow noise measurements. Then the electrical output of the sound level meter in response to a 1000 Hz calibration tone at 94 dB SPL generated by a calibrator (Quest 211A) was recorded. The levels of this tone at the input ranges of 30–100 and 70–140 dB SPL served as the 94 dB references for the recordings made at the corresponding input ranges.

E. Data analysis

The overall and one-third octave band levels from 100 to 8000 Hz of the flow noise recordings were analyzed using MATLAB. The correlation coefficients and average differences between the levels in different data sets were analyzed. Correlation coefficients were used to infer similarities between the hearing aids' microphone sensitivity to flow noise at different head angles. They were calculated using the directivity data from 0° to 350°. High correlation coefficients suggested that flow noise levels in the two data sets rise and fall at corresponding head angles, although their magnitude might differ by a relatively constant number.

Average differences were calculated by subtracting the levels of one data set from the corresponding levels at the same head angles of another data set, taking the absolute values of the differences, and then calculating the mean of the differences. The average differences calculated using the absolute values of the differences were generally higher than simple differences because they treated levels lower or higher than the reference (i.e., negative and positive differences, respectively) as variations in the measurements.

In general, a high correlation coefficient with low average differences indicated that the polar patterns of the two data sets had similar trends/polar shapes and levels. A high correlation coefficient with high average differences indicated that the polar patterns of the two data sets had similar trends/shapes but their levels differed by an amount close to the average difference. A low correlation coefficient indicated that either the shapes of the polar patterns were grossly different or the levels in one data set were sometimes greater than the other set and other times vice versa.

III. RESULTS

The 125, 500, and 2000 Hz one-third octave band levels are reported along with overall levels to represent flow noise levels at very-low-, low-, and high-frequency regions, respectively. Note that the limiters in the hearing aid made the polar pattern look circular at high velocities, especially at frontal or rearward incidences. The actual level at the microphone output could be much higher.

Angles are reported in clockwise direction and the polar patterns are divided into four quadrants by ±45° oblique lines. Frontal incidence refers to angles between 320° and 40°, rearward incidence refers to angles between 130° and 220°, and right- or left-lateral incidence refers to angles between 230° and 310° or between 50° and 120°, respectively. In addition, hearing aid facing upstream generally refers to the angles between 230° and 310° when the hearing aid was worn on the right ear and angles between 50° and 120° when the hearing aid was on the left ear, and vice versa for hearing aid facing downstream.

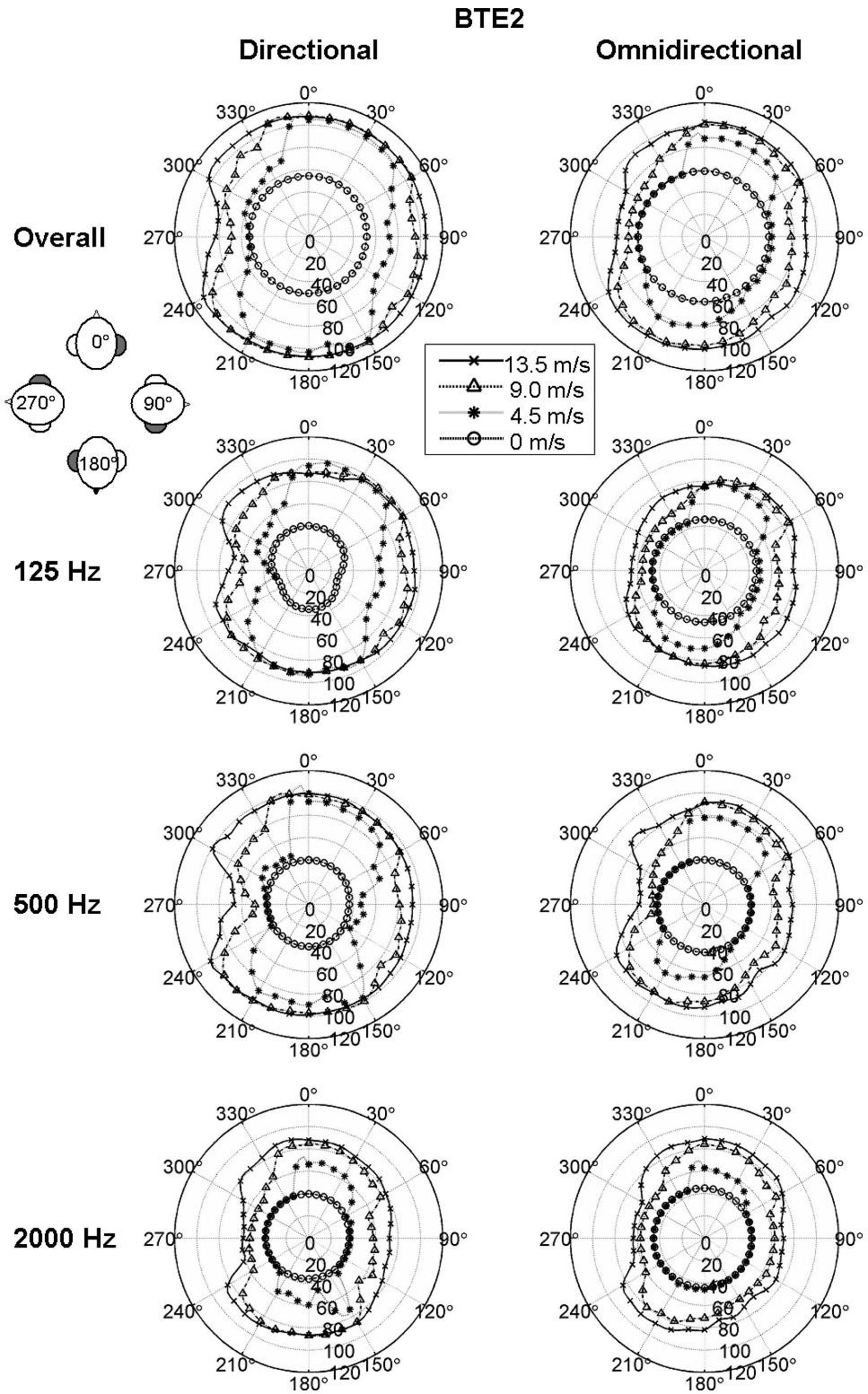


FIG. 6. Overall and one-third octave band levels of flow noise measured at the output of BTE2, which was set to the directional (DIR) and omnidirectional (OMNI) modes.

A. Flow noise responses

Figures 5 and 6 show the polar patterns of the noise levels at the outputs of BTE1 and BTE2 when both were worn on the right ear. Similar maximum values over several angles (e.g., BTE1 OMNI between 0° and 50° at 13.5 m/s in the overall level plot in Fig. 5) indicate that the hearing aid output has reached its maximum level as imposed by the input and/or output limiters in the signal path. Additionally, if the noise levels at 4.5 m/s were similar to those in the absence of a flow (i.e., ambient noise level at 0 m/s), it is likely that the flow noise was below the ambient noise level at this angle.

TABLE I. The correlation coefficients and average differences between BTE1 and BTE2 at three flow velocities when both hearing aids were worn on KEMAR's right ear. All correlation coefficients were statistically significant. The asterisks (*) mark the entries calculated from data set when overall levels reached the maximum output levels at least at three angles and the carets ($^\wedge$) mark the entries calculated from data set when flow noise levels fell below the ambient noise levels at least at three angles.

	Correlation coefficient			Average difference (dB)		
	4.5 m/s	9.0 m/s	13.5 m/s	4.5 m/s	9.0 m/s	13.5 m/s
			Overall level			
DIR	0.87$^\wedge$	0.78*	0.71*	12.6$^\wedge$	8.4*	6.7*
OMNI	0.93$^\wedge$	0.89	0.83*	2.8$^\wedge$	4.2	5.4*
			125 Hz third octave band level			
DIR	0.88$^\wedge$	0.84*	0.76*	14.5$^\wedge$	9.7*	6.7*
OMNI	0.92$^\wedge$	0.92	0.90*	5.3$^\wedge$	3.5	3.3*
			500 Hz third octave band level			
DIR	0.87$^\wedge$	0.75*	0.66*	9.2$^\wedge$	8.0*	7.1*
OMNI	0.93$^\wedge$	0.87	0.76*	3.9$^\wedge$	6.4	7.1*
			2000 Hz third octave band level			
DIR	0.88$^\wedge$	0.88*	0.81*	8.7$^\wedge$	5.2*	7.6*
OMNI	0.95$^\wedge$	0.82	0.72*	1.9$^\wedge$	4.2	6.9*

Table I summarizes the correlation coefficients and average differences between the two hearing aids. Coefficients calculated with levels below the ambient noise levels are marked with a #. Coefficients for which either or both hearing aids reached their maximum output level at more than three angles are marked with an *. All the correlation coefficients reported were statistically significant at $p < 0.01$ ($df = 34$).

Before the levels reached the maximum output levels (i.e., DIR at 4.5 m/s and OMNI at 4.5 and 9.0 m/s), the correlation coefficients between BTE1 and BTE2 were generally very high for corresponding microphone modes (between 0.82 and 0.95). This indicated similar BTE1 and BTE2 polar patterns, even though the flow noise fell below the ambient noise levels at some angles. The correlation coefficients for polar patterns calculated with data reaching the maximum output levels were somewhat lower (between 0.66 and 0.88). DIR at 13.5 m/s was generally the lowest, which was likely because the overall levels of BTE2 were limited to the same values at most angles but those of BTE1 showed level variations.

The average differences between BTE1 and BTE2 OMNI were between 1.9 and 7.1 dB, which were generally lower than the average difference between DIR (i.e., 5.2–14.5 dB), suggesting that DIR of different hearing aids had higher level variations than OMNI in wind noise.

1. Similarities between BTE1 and BTE2

In general, BTE1 and BTE2 polar patterns were similar (Figs. 5 and 6), which suggests that the two hearing aids were subjected to similar aerodynamic conditions due to similar microphone locations. Specifically, we observe the following:

(i) The polar patterns of the ambient tunnel noise levels at 0 m/s looked circular except that of BTE2 DIR at 125 Hz.

(ii) At high flow velocities, the flow noise level reached the maximum output level of the hearing aids (i.e., BTE1 DIR=114 dB, OMNI=109 dB; and BTE2 DIR=108 dB, OMNI=102 dB).

(iii) The flow noise levels were generally the lowest when the hearing aids were facing upstream. The flow noise levels at 4.5 m/s could drop below the ambient tunnel noise levels at 0 m/s at these angles.

(iv) The flow noise levels were generally the highest for frontal or rearward incidences. They were the highest between 350° and 60° and between 160° and 220° at 4.5 m/s. The difference between the lowest and highest overall levels within one 360° run could be as high as 53 dB (i.e., BTE2 DIR at 10 m/s). As the flow velocity increased, the flow noise level increased and the angles with high flow noise levels also broadened (e.g., the overall level of BTE1 DIR had high levels between 350° and 50° at 4.5 m/s and it had high levels between 330° and 50° at 9.0 m/s).

(v) When the hearing aids were facing downstream, the flow noise levels were generally higher than when the hearing aids were facing upstream and lower than the levels for frontal or rearward incidences.

2. Differences between BTE1 and BTE2

Several differences were also observed.

(i) As the maximum overall output level of BTE1 was slightly higher than that of BTE2, the overall levels of BTE2 DIR were limited to the maximum output level at most of the head angles (from 300° to 240° clock-

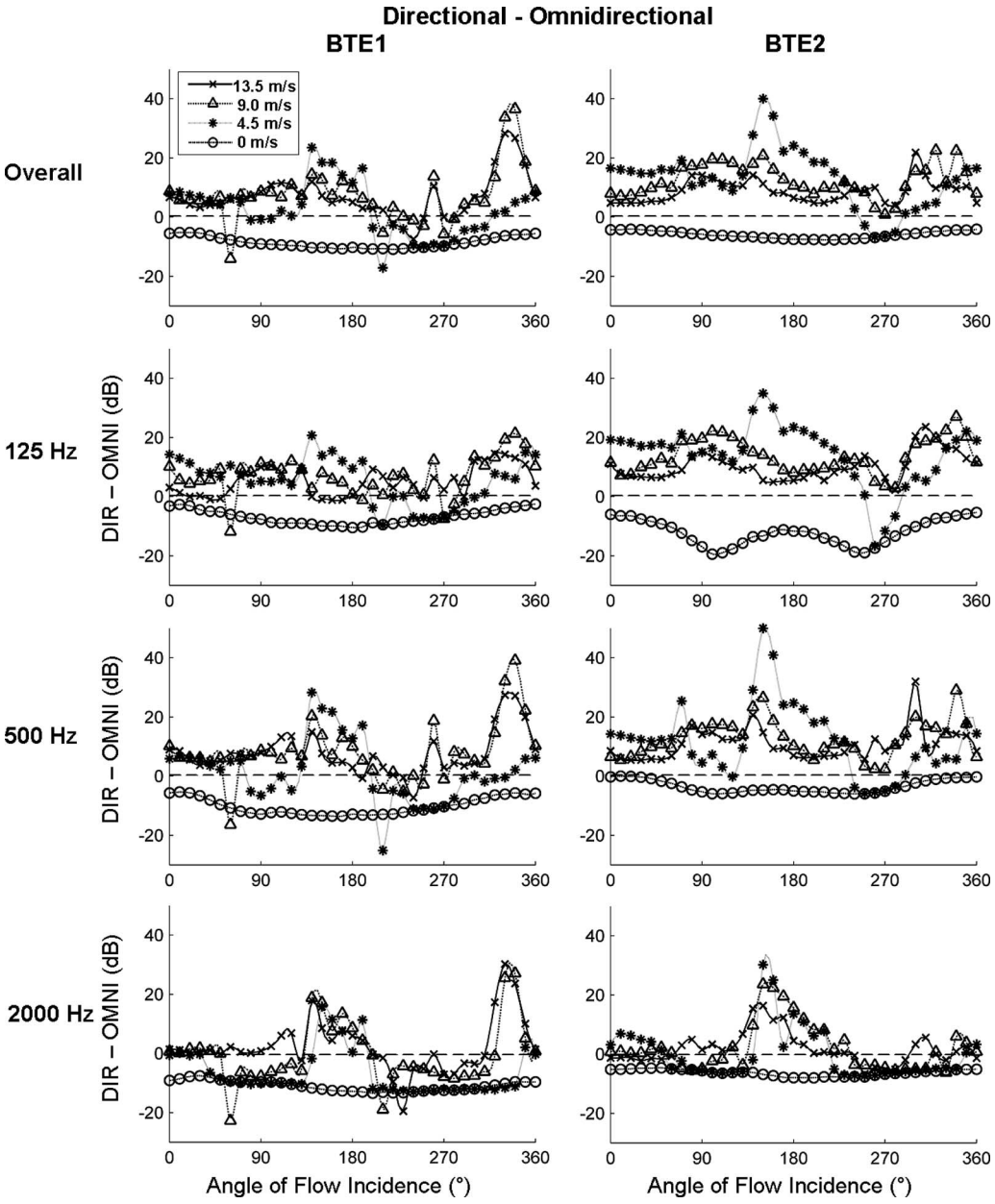


FIG. 7. Overall and one-third octave band level differences calculated by subtracting the noise level of the omnidirectional microphone (OMNI) from its corresponding directional microphone (DIR).

wise) whereas BTE1 DIR only reached its maximum output level at the frontal and rear incidence angles.

(ii) The overall plots of BTE1 DIR had a sudden level increase at 260° whereas BTE2 had a slight level increase between 300° and 310°.

(iii) In the 125 Hz plot between 0° and 10°, BTE1 DIR had similar levels from 4.5 to 13.5 m/s whereas BTE2 DIR yielded higher levels at 4.5 m/s than at 9.0 and 13.5 m/s.

B. DIR and OMNI comparison

The level differences of DIR minus OMNI for BTE1 and BTE2 are shown in Fig. 7. Table II also shows the correlation coefficients and average differences between the DIR and OMNI modes of the hearing aids. The data sets with the levels at three or less angles reaching the maximum output level generally had high to very high correlation coefficients (0.76–0.92), indicating that the polar patterns of DIR and OMNI had similar shapes. The exception is that the correlation coefficient at BTE2 at 4.5 m/s at 2000 Hz was slightly lower (i.e., 0.68). This was probably because the flow noise levels of BTE2 DIR were much higher than the ambient noise level at 0 m/s at the rearward incidence angles and showed level variations whereas those of BTE2 OMNI dropped below the ambient noise level at those angles and showed minimal level variations. Although similar observations are also noticed in BTE1 at 2000 Hz one-third octave band, the flow noise levels of BTE1 DIR for rearward incidence angles were relatively lower than those for BTE2 DIR. The correlation coefficient between BTE1 DIR and OMNI was, therefore, not lowered considerably.

The correlation coefficients were also high (between

TABLE II. The correlation coefficients and average differences between DIR and OMNI at three flow velocities when both BTE1 and BTE2 were worn on KEMAR's right ear. All correlation coefficients were statistically significant. The * marks the entries calculated from data set when overall levels reached the maximum output levels at least at three angles and the ^ marks the entries calculated from data set when flow noise levels fell below the ambient noise levels at least at three angles.

	Correlation coefficient			Average difference (dB)		
	4.5 m/s	9.0 m/s	13.5 m/s	4.5 m/s	9.0 m/s	13.5 m/s
			Overall level			
BTE1	0.89^	0.82	0.81*	7.3^	9.1	7.6*
BTE2	0.86^	0.89*	0.86*	14.1^	12.6*	8.9*
			125 Hz third octave band level			
BTE1	0.92^	0.84	0.85*	7.9^	8.1	5.5*
BTE2	0.87^	0.84*	0.76*	16.3^	13.8*	10.2*
			500 Hz third octave band level			
BTE1	0.87^	0.83	0.84*	8.0^	9.6	8.1*
BTE2	0.85^	0.90*	0.84*	13.2^	12.9*	10.9*
			2000 Hz third octave band level			
BTE1	0.76^	0.76	0.80*	9.1^	7.9	6.1*
BTE2	0.68^	0.82*	0.83*	6.8^	5.9*	3.8*

0.76 and 0.90) for data sets with levels limited at a greater number of angles (i.e., BTE1 at 13.5 m/s and BTE2 at 9.0 and 13.5 m/s). This could be because DIR and OMNI of the same hearing aid shared more commonalities in signal processing algorithm settings and physical characteristics.

The following observations were made for the DIR and OMNI modes of the two behind-the-ear hearing aids.

(i) In the absence of a flow (0 m/s), DIR generally had lower levels than OMNI (Figs. 5 and 6). This difference can also be seen as the negative lines in Fig. 7, which shows the difference of DIR minus OMNI.

(ii) DIR has higher levels than OMNI in most angles and flow velocities (i.e., positive values in Fig. 7). The average differences ranged from 3.8 to 16.3 dB (Table II), but they may be as high as 35–45 dB in some angles (Fig. 7).

(iii) DIR tended to reach the maximum overall levels at more angles and at lower flow velocities than OMNI (Figs. 5 and 6). The flow noise at angles between 300° and 350° and between 140° and 180° was always higher in DIR than in OMNI at flow velocities of 9.0 and 13.5 m/s.

(iv) In some angles, the flow noise of DIR was lower than that of OMNI (i.e., negative values in Fig. 7), especially at 2000 Hz. For angles at which the flow noise was equal to or below the ambient noise levels, the differences between DIR and OMNI at 4.5 m/s equaled those at 0 m/s (e.g., comparing the levels at 2000 Hz one-third octave band in Figs. 5 and 6 with Fig. 7 for angles around 270°).

(v) Although the output levels of both modes were programed to be as close as possible using pink noise during the fitting process, DIR tended to have a higher maximum overall level than OMNI for both BTE1 and BTE2 (i.e., the arc of similar values in the overall level plots had a higher level in DIR than in OMNI plots). One of the reasons could be because both of the omnidirectional microphones in DIR were saturated whereas only one omnidirectional microphone in OMNI was saturated.

C. Repeatability

Figure 8(A) shows overall levels of BTE2 DIR (Run2) recorded immediately after the BTE2 DIR (Run1, plots shown in Fig. 6). The correlation coefficients and average differences for the overall and one-third octave band levels between the two measurements are also tabulated in Table III.

The correlation coefficients between Run1 and Run2 of BTE2 DIR ranged between 0.93 and 1.0 and the average differences were within 2.3 dB at the overall and the one-third octave band levels at all flow velocities, indicating that the flow noise measurement was highly repeatable.

D. Variability

Figure 8(B) shows the Cartesian plot of the overall flow noise levels of BTE1 DIR measured 14 days apart at flow velocities of 9.0 and 0 m/s and Fig. 8(C) shows that of BTE1 OMNI measured at 13.5 and 4.5 m/s. Additionally, the correlation coefficients and average differences between Day 1 and Day 2 for overall and one-third octave band levels are shown in Table III.

The very high correlation coefficients (>0.92) and very low average differences (<4.0 dB) suggested that the flow noise measurements were highly repeatable within a 14-day period (Table III). Obvious differences were only seen at 210° for BTE1 OMNI and at 60°, 130°, 260°, and 330° for

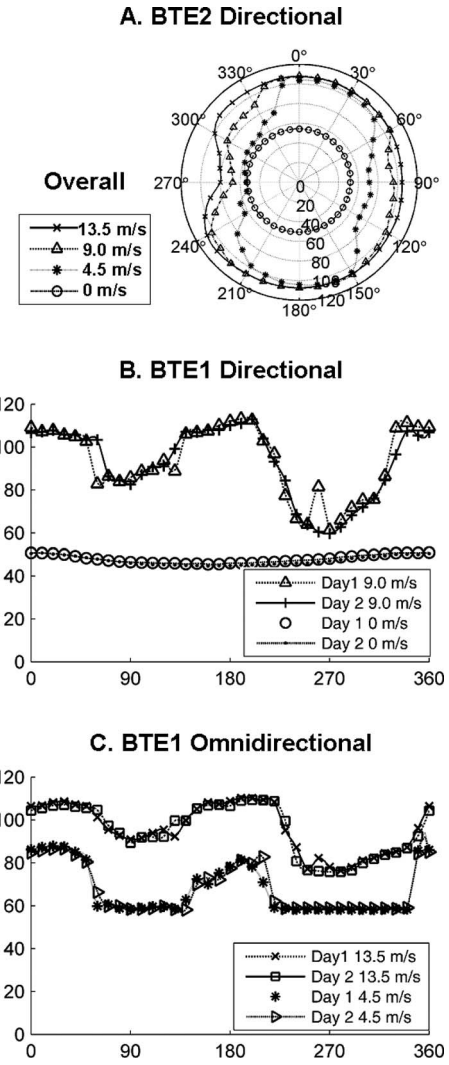


FIG. 8. (A) Overall of BTE2 DIR Run2 taken immediately after Run1 without dissembling the wind tunnel setup and the recording equipment. Overall levels of BTE1 (B) DIR and (C) OMNI taken 14 days apart.

BTE1 DIR. These differences were likely due to momentary fluctuations. Sometimes, the levels could shift by 10° at angles bordering a sharp change in flow pattern [e.g., 60° at 9.0 m/s in Fig. 8(B)].

E. Left vs right comparisons

Figure 9 shows the polar patterns of overall flow noise levels at the output of BTE1 when it was worn on KEMAR's left ear (BTE1_Left). The correlation coefficients and average differences between BTE1 and BTE1_Left were tabulated in Table III. They were calculated by inverting the flow noise levels recorded from 0°–350° to 350°–0° for the BTE1_Left data (i.e., mirror images).

Very high correlation coefficients (between 0.89 and 0.96) were found between the two sets of measurements, indicating that they had similar shapes when compared in mirror image. The average differences ranged between 2.3 and 5.5 dB, which were slightly higher than the repeated measures of BTE1 discussed above. This might be in part due to KEMAR's asymmetric facial features.

The measurements at 18.5 and 22.5 m/s indicated that flow noise at the hearing aid output was limited to its maximum output level at higher velocities, especially for frontal or rearward incidence. The flow noise levels when the hearing aid was facing upstream or downstream were consistently lower than those at other angles.

F. Flow around cylinder

Figure 10 shows flow noise recorded by the B&K microphone mounted on the cylinder at four flow velocities. Caution is needed in interpreting the measurement data. Flow noise levels cannot be directly compared with those recorded at the hearing aid output because KEMAR's cross section was oblong whereas the cylinder was circular with a relatively larger diameter. In addition, the size of the B&K microphone diaphragm was much larger than the hearing aid microphone diaphragm.

One prominent similarity between the hearing aid and cylinder measurements was that the lowest flow noise was recorded when the microphone was facing upstream. This was due to the presence of a large region of separated flow. Several differences, however, were also found.

(i) The highest levels for the cylinder measurement occurred when at angles that were close to 90° (i.e., the microphone diaphragm was facing downstream) at 4.5 m/s and between 20° and 170° at 9.0, 13.5, and 18.5 m/s. The highest levels for hearing aids, however, occurred when KEMAR was facing upstream or downstream (i.e., the hearing aid microphone was on the sides).

(ii) As flow noise increased, the levels of the cylinder measurements increased monotonically in most angles, suggesting that the cessation of level increase in the hearing aid measurements was indeed due to the limiters in the signal path imposing a maximum output level. The exceptions to the monotonic increase in levels were probably due to fluctuations in flow noise during the recording.

(iii) The overall and 125 Hz one-third octave band ambient tunnel noise levels at 0 m/s were generally higher in the cylinder than the hearing aid measurements. This was likely due to the relatively flat frequency response of the B&K microphone at low frequencies whereas those of hearing aids often have a roll-off below 200 Hz.

IV. DISCUSSION

In this study, we recorded flow noise at the output of two behind-the-ear digital hearing aids that did not have any wind noise management or cancellation algorithms. The following discuss in detail the factors that affected flow noise at the hearing aid output.

A. Level limiters

Recall that digital hearing aids may have several points in the signal path to limit the hearing aid output of high-level sounds. In this study, the effects of input and/or output level limiters could be observed in the polar patterns in several ways. First, flow noise at high velocity could saturate the

TABLE III. Correlation coefficients and average differences for BTE1 and BTE2 at three flow velocities. All the correlation coefficients were statistically significant ($p<0.01$).

	Correlation coefficient			Average difference (dB)		
	4.5 m/s	9.0 m/s	13.5 m/s	4.5 m/s	9.0 m/s	13.5 m/s
			Overall level			
BTE2 DIR Run1 vs Run2	0.97	0.98	0.99	1.4	1.6	0.6
BTE1 DIR Day 1 vs Day 2	–	0.93	–	–	3.4	–
BTE1 OMNI Day 1 vs Day 2	0.97	–	0.98	1.5	–	1.8
BTE1 vs BTE1_wornL DIR	0.94	0.96	0.95	4.5	3.5	2.6
BTE1 vs BTE1_wornL OMNI	0.92	0.95	0.95	3.3	3.8	3.3
			125 Hz third octave band level			
BTE2 DIR Run1 vs Run2	1.00	0.98	0.99	1.3	1.4	0.7
BTE1 DIR Day 1 vs Day 2	–	0.90	–	–	3.3	–
BTE1 OMNI Day 1 vs Day 2	0.97	–	0.98	1.6	–	1.8
BTE1 vs BTE1_wornL DIR	0.94	0.95	0.89	4.8	3.8	3.2
BTE1 vs BTE1_wornL OMNI	0.94	0.95	0.95	3.7	3.8	2.7
			500 Hz third octave band level			
BTE2 DIR Run1 vs Run2	0.98	0.98	0.91	2.3	2.0	1.3
BTE1 DIR Day 1 vs Day 2	–	0.92	–	–	4.0	–
BTE1 OMNI Day 1 vs Day 2	0.97	–	0.97	1.9	–	2.4
BTE1 vs BTE1_wornL DIR	0.91	0.96	0.94	5.5	4.0	3.0
BTE1 vs BTE1_wornL OMNI	0.89	0.95	0.93	4.3	4.1	3.9
			2000 Hz third octave band level			
BTE2 DIR Run1 vs Run2	0.93	0.98	0.98	1.9	1.5	1.2
BTE1 DIR Day 1 vs Day 2	–	0.95	–	–	2.5	–
BTE1 OMNI Day 1 vs Day 2	0.99	–	0.99	1.1	–	1.6
BTE1 vs BTE1_wornL DIR	0.92	0.96	0.91	3.3	3.6	5.0
BTE1 vs BTE1_wornL OMNI	0.97	0.91	0.93	2.3	3.8	4.6

microphone, resulting in digital distortions that could increase the level of high-frequency components at the hearing aid output. Second, the limiters in hearing aids could make the noise levels at two or more flow velocities similar (e.g., overall plots of BTE1_Left DIR at 13.5, 18.5, and 22.5 m/s). Third, the limiters could impose a limit on the maximum output level resulting in similar maximum values at high flow velocities, making the polar patterns at the frontal or rearward incidences look circular.

Fourth, the limiters could created a counter-intuitive phenomenon in the polar patterns—flow noise at lower flow velocities having a higher level than flow noise at higher velocities at low frequencies. For example, in the 125 Hz one-third octave band at angles between 0° and 10°, the levels of BTE2 DIR at 4.5 m/s were higher than those at 9.0 m/s (Fig. 6). Flow noise has a high level of low-frequency components at the microphone output. As the flow velocity increased, the high-frequency components of the flow noise increased. When the signal reached the broadband limiting algorithm, more gain reduction was provided to the signals recorded at a higher flow velocity than at a lower velocity. As the overall level was limited to the same/similar maximum output level, the levels of the low-frequency region for a signal with higher levels of high-frequency contents (i.e., flow noise at a higher velocity) was estimated to be lower than the levels of the low-frequency region for a signal with lower levels of high-frequency contents (i.e., flow noise at a lower velocity).

Another possibility was that when the microphone output reached the limiters implemented in different frequency channels (channel-limiters), the strong low-frequency components were limited by the limiters at low-frequency channels. As flow velocity was increased, the channel-limiter continued to act on the low-frequency components resulting in a relatively constant level at the low frequencies but allowing the high-frequency components to continue to increase in level. As the signal reached the broadband output limiter, flow noise generated at higher flow velocities incurred more gain reduction than that generated at lower flow velocities. The combination effect of the channel- and broadband-limiters resulted in lower low-frequency component levels at higher flow velocities than at lower flow velocities. Future studies are needed to study the flow noise patterns without the constraints of the limiters.

B. Influence of head angle

Flow noise levels recorded at frontal and rearward incidence angles were generally higher than those recorded when the hearing aid was facing downstream, which, in turn, was higher than when the hearing aid was facing upstream. These

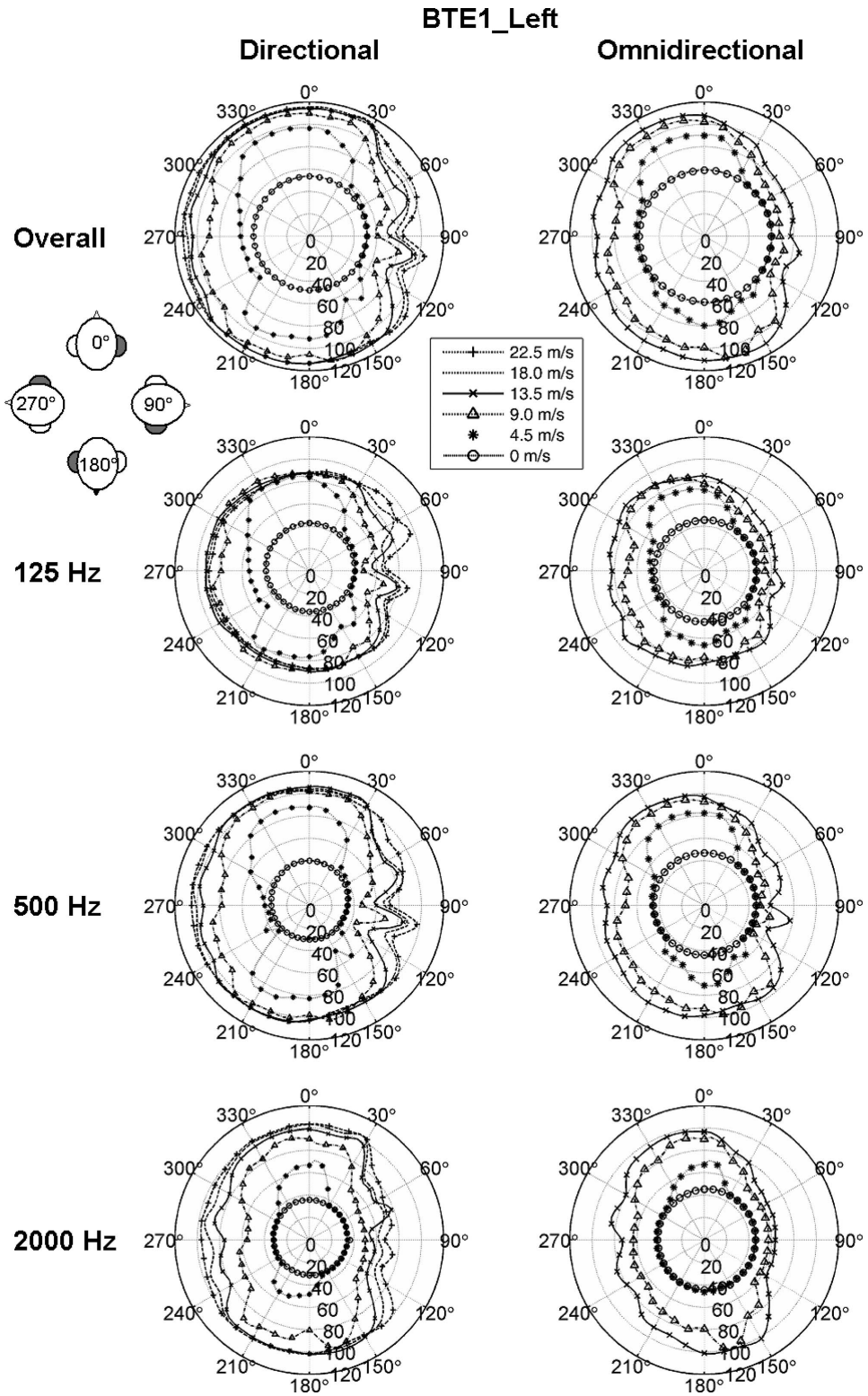


FIG. 9. Overall and one-third octave band levels of flow noise measured at the output of BTE1 from different head angles when it was worn on KEMAR's left ear.

results were consistent with those reported by Beard and Nepomuceno (2001) and Thompson and Dillon (2002). As the flow velocity increased, the noise level at frontal and rearward incidences reached the maximum output level first and then other angles followed. To shed light on the mechanisms of this head-angle-dependent phenomenon, we examined the aerodynamic patterns when a flow passes by a cylindrical body.

For a laminar flow around a cylindrical body, the flow accelerates and reaches a maximum velocity on the lateral sides of the cylinder (Fig. 11). An incompressible fluid is assumed at the relatively low flow velocities dealt with in this study. At a very low flow velocity [Fig. 11(A)], the flow pattern is characteristically laminar both upstream and downstream (i.e., the flow layers do not mix). The air bends outward on approaching the body, goes around it, and bends

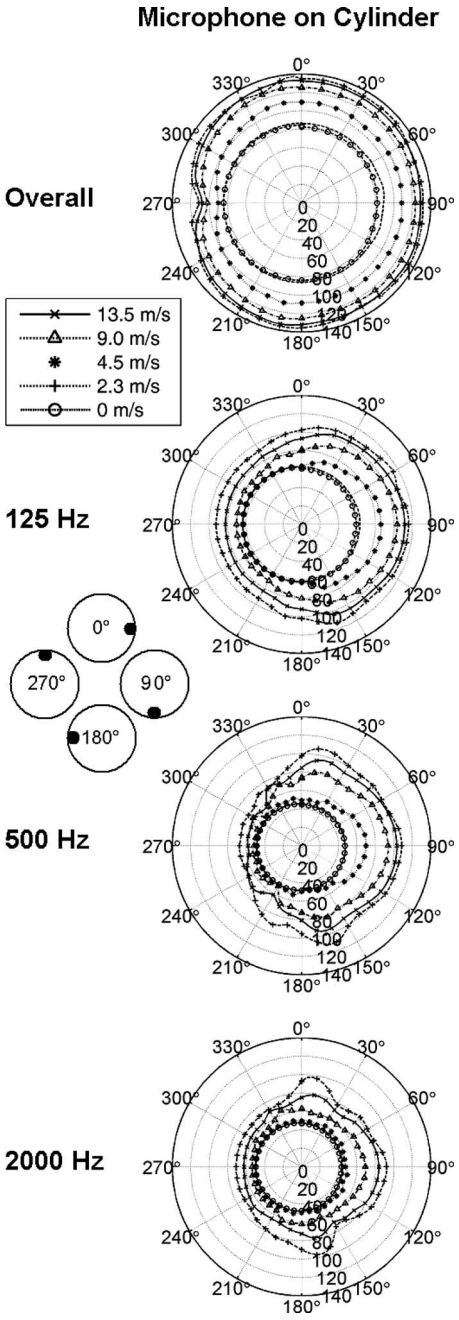


FIG. 10. Overall and one-third octave band levels of flow noise measured at the output of a B&K microphone when it was mounted on a cylinder with 15-cm diameter.

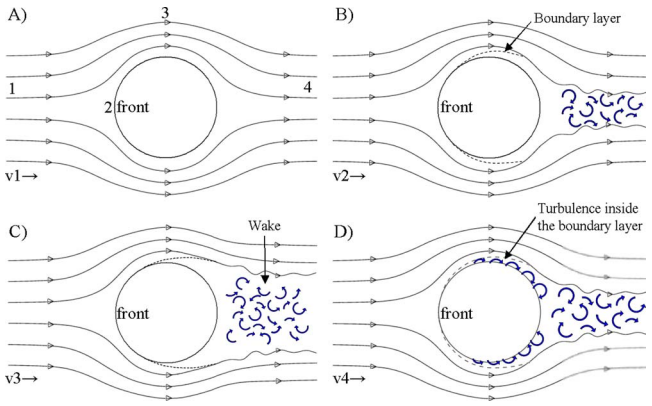


FIG. 11. (Color online) Aerodynamics of a flow passing a body at various flow velocities ($v1<v2<v3<v4$). (A) At a low flow velocity, $v1$, laminar flow is observed both upstream and downstream. Point 1: upstream flow; point 2: stagnation point; point 3: lateral side of the body; and point 4: downstream flow. (B) At a higher flow velocity, $v2$, the flow becomes turbulent in the wake. A thin boundary layer is formed around the sides of the body and vortices are formed in the wake. Note: The thickness of the boundary layer is exaggerated for clarity. (C) At flow velocity higher than $v2$, the wake becomes larger. (D) Above a certain critical velocity, $v4$, turbulence is formed inside the boundary layer and the size of the wake is reduced.

inward in the wake. Therefore, the flow velocity of the streamwise component on the lateral sides of the body (point 3) needs to be greater than that upstream (point 1) in order for the volumetric flow rate to be constant along the streamwise direction at the back of the body (point 4). In an idealized condition, a stagnation point is present at the center of the front surface (point 2). The flow velocity at this point is theoretically 0 m/s.

At higher flow velocities [Figs. 11(B) and 11(C)], viscosity causes flow velocity on the surface of the body to be still, forming a thin boundary layer. The streamlines beyond the boundary layer are free-flowing. The air flow velocity inside the boundary layer increases away from the surface of the body, but it is lower than the outer air streams. Beyond a certain point on the lateral sides of the cylinder, the air inside the boundary layer stops and reverses its flow direction as it encounters high pressure in the wake region. This flow reversal leads to the separation of air flow from the free-flowing stream and creates eddies in the wake. The size of the eddying wake increases with the flow velocity [i.e., the turbulent wake in Fig. 11(C) is greater than that in Fig. 11(B)]. Above a critical flow velocity [Fig. 11(D)], which depends on the size and surface characteristics of the body and the viscosity and density of the air, turbulence is formed inside the boundary layer. Above this velocity, the point of separation moves further downstream and the size of the wake is reduced (Anderson, 2005; Yip, 2007).

Relating the flow aerodynamics in Fig. 11 to the microphone-on-cylinder measurements in Fig. 10, it appeared that as the flow velocity increased from 0 to 2.3 m/s, the flow noise level at 125 Hz one-third octave band increased slightly when the microphone was facing downstream (i.e., between 60° and 120°) compared to facing upstream. The level increases were likely due to turbulent eddies in the wake of the cylinder. As the flow velocity increased to 4.5 m/s, flow noise levels in the 125 and 500 Hz one-third octave bands increased in level and spread to wider angles in the wake region, which was consistent with the increase in wake size as flow velocity increased [Figs. 11(B) and 11(C)]. As the flow velocity further increased to 9.0, 13.5, and 18.5 m/s, the flow noise level increased further in the wake and at all angles. Flow noise at higher frequencies also became prominent. Part of this level increase could be due to the increase in ambient tunnel noise level at higher flow velocities. Part of it would be due to the increase in small scale turbulence.

The unusual level increase at angles between 0° and 20° and between 160° and 180° at flow velocities of 9.0–18.5 m/s in the 500 and 2000 Hz one-third octave bands suggests

that factors other than turbulence might have contributed to the flow noise at the microphone output. From Fig. 11, the highest flow velocities are reached on the sides of the body. At the critical velocities, turbulent eddies were formed inside the boundary layer [Fig. 11(D)]. It was possible that the combination of high velocity and turbulence caused the sudden increase in flow noise levels. This interpretation is consistent with the observation that flow noise levels increased with the flow velocity. In the case of KEMAR, the flow might be tripped by the presence of the pinnae or other protuberances from the sides of the head, causing separation and increased generation of turbulence. Combined with the fact that the hearing aids were at the region with the highest flow velocity, the flow noise at frontal and rearward incidence angles, therefore, became much higher than the lateral incidence angles.

Consistent with the observation that the lowest flow noise levels were generated when the hearing aid microphone was facing upstream (Figs. 5 and 6), the lowest overall level for the microphone-on-cylinder measurement also occurred when the microphone was directly facing upstream (i.e., 270° in Fig. 10). This was likely because both flow velocity and turbulence were the lowest at the "stagnation" point. The increase in level could be partly due to the increase in ambient tunnel noise in the wind tunnel (Fig. 2) and partly due to the microphone was larger than an infinitely small point, which was assumed in an idealized case.

When the microphones were in the wake, flow noise levels at the hearing aid output were relatively lower than those at the corresponding angle recorded on the cylinder. This could partly be because KEMAR's head was oblong. The hearing aid microphone was, therefore, further away from the turbulence in the wake than the microphone mounted on the cylinder. Future studies utilizing flow visualization techniques are needed to study the relationship between anatomic structures, flow characteristics, and flow noise in hearing aids.

C. Influence of microphone type

1. No flow

In the absence of a flow (i.e., 0 m/s), the directional microphones of BTE1 and BTE2 provided more attenuation to far-field ambient tunnel noise and resulted in lower levels at the hearing aid outputs. This noise reduction effect is shown as the negative OMNI-DIR differences at 0 m/s in Fig. 7. The average amounts of attenuation in the overall, 125, 500, and 2000 Hz were 8.6, 7.0, 10.5, and 11.0 dB for BTE1, respectively, and 5.9, 12.2, 3.6, and 6.3 dB for BTE2, respectively.

The level variations at 0 m/s shown in Figs. 5 and 6 were likely a combination effect of the head shadow, the sensitivity/directivity pattern of the microphone, and the actions of the expansion algorithm. The polar patterns of OMNI were generally circular with slightly higher levels when the hearing aid was facing the wind tunnel inlet (i.e., around 270° in Figs. 5 and 6 and around 90° in Fig. 9). This suggested that the ambient noise was mainly from the inlet instead of the outlet side of the wind tunnel in the absence of a flow, and the head shadow effect resulted in lower ambient noise levels when the microphone was not facing the wind tunnel inlet.

On the other hand, the polar patterns of DIR showed relatively larger angle-to-angle level variations (Figs. 5 and 6). Similar to the data reported for OMNI, the levels of DIR at 270° were slightly higher than those at 90° due to the head shadow effect. In addition, the polar patterns for DIR showed a lower level for 180° compared to 0°, which was likely because DIR had lower sensitivity to sounds coming from the back. In general, the polar patterns of DIR measured in the wind tunnel lacked well defined nulls compared to those measured in anechoic chamber. Only BTE2 DIR at 125 Hz one-third octave band showed lower levels at the expected nulls of hypercardioid microphones (i.e., around 135° and 225°).

Previous studies suggested that directional microphones process sounds within and outside of the critical distance differentially (Leeuw and Dreschler, 1991; Ricketts and Hornsby, 2003). Directional microphones provide an equal amount of noise attenuation for sounds coming from the reverberant field (i.e., outside the critical distance). When signals are within the critical distance, however, they provide different amounts of attenuation to sounds coming from different directions and the amount of attenuation approaches the polar patterns measured in an anechoic chamber. Our results suggested that the sources of the ambient noise were in the reverberant field of the directional microphones at most frequency regions. The more distinct shape at 125 Hz for BTE2 DIR was speculated to be because the noise source was in the borderline of the critical distance at the 125 Hz one-third octave band.

Finally, Fig. 7 shows an unusually large amount of noise attenuation provided by DIR compared to OMNI. This was likely because the directional microphones attenuated the noise levels to below the expansion thresholds at the low-frequency channel of the hearing aids. As the expansion algorithms are designed to provide less gain to low-level signals than high-level signals, the noise at the DIR output was therefore reduced and yielded a very low level at the hearing aid output.

2. With flow

In the presence of a flow, DIR had a higher sensitivity to near-field flow variations and they generated higher flow noise levels than OMNI in most head angles (i.e., positive differences in Fig. 7). At some angles, the directional microphones were found to have lower outputs than the omnidirectional microphones (i.e., negative differences in Fig. 7). The following discusses several possible causes for such observations.

First, when there is a considerable change in flow noise between adjacent angles, OMNI may appear to have a higher flow noise level than DIR (e.g., at 60° of BTE1 in Fig. 5). The variability of BTE1 DIR in Fig. 8(B) shows that flow noise level at 60° varied from just over 80 dB SPL in Day 1 to over 110 dB SPL in Day 2. While this level difference seemed to be very large, this is likely a result of momentary fluctuations because the level at 60° had similar level as 70°

on Day 1 but it had similar level 50° on Day 2. When events like this happened in either DIR or OMNI measurements, DIR may appear to have lower flow noise levels than OMNI.

Second, when DIR reduced the ambient noise in the far field in the wind tunnel, DIR minus OMNI would yield a negative difference (Fig. 7). For example, BTE1 OMNI had a higher overall noise level than BTE1 DIR at 270° at 0 m/s in Fig. 5. As the flow noise at 4.5 m/s for both DIR and OMNI was lower than their corresponding ambient tunnel noise levels, the level recorded at the hearing aid output equaled the ambient noise level. The overall DIR-OMNI difference in Fig. 7, therefore, shows a negative difference at 270° because DIR reduced the ambient noise level.

Third, a negative DIR-OMNI difference could also be due to the attenuation of the wind tunnel ambient noise by DIR at higher flow velocities. For example, BTE1 DIR-OMNI difference at 2000 Hz is negative at 270° at 9.0 m/s in Fig. 7. Given that the BTE1 DIR-OMNI difference was −12 dB at 270° at 2000 Hz at 0 m/s, the flow noise of DIR needed to be at least 12 dB higher than that of OMNI in order to yield a positive DIR-OMNI difference. Our measurements, however, indicated that BTE1 OMNI and DIR were estimated to be 57 and 49.5 dB at 270° in the 2000 Hz plot at 9.0 m/s, respectively, resulting in a difference of −8.5 dB.

Fourth, the limiters that are responsible for setting the maximum output levels of the hearing aids can potentially diminish the DIR-OMNI difference or turn it negative. For example, the overall level of BTE2 between 0° and 50° exhibited a decrease in DIR-OMNI difference as the flow velocity increased from 4.5 and 13.5 m/s (Fig. 7). Figure 6 shows that the overall levels of BTE2 DIR were much higher than OMNI at 4.5 m/s (i.e., difference equaled 15–16 dB in Fig. 7). At 9.0 m/s, both BTE2 DIR and OMNI approached their maximum output level, resulting in lesser DIR-OMNI differences (i.e., 7–11 dB) than at 4.5 m/s. At 13.5 m/s, both BTE2 DIR and OMNI reached their own maximum output levels (Fig. 7) and the DIR-OMNI difference was reduced to the difference between the maximum output levels of DIR and OMNI (i.e., approximately 5 dB). Theoretically, if BTE2 DIR had the same or lower maximum output level than OMNI, the differences would have disappeared or become negative, respectively.

D. Summary and applications

By recording flow noise outputs from commercially available digital hearing aids and taking the standard signal processing algorithms into account, this study provided realistic flow noise data useful for clinical applications. The results indicated that both directional and omnidirectional microphones yielded the lowest flow noise levels when the microphones were facing directly upstream. Relatively low levels were also generated when the microphones were facing downstream. If unilateral hearing aid or cochlear implant users or users with a "good" and a "poor" ear want to conduct conversations in windy environments, it is advisable to position the "good" ear facing upstream and the "poor" ear facing downstream. For bilateral hearing instrument users including hearing protector users, it is best to position themselves so that one hearing aid faces upstream and the other faces downstream.

Our results also revealed that the limiters in the signal path can reduce the hearing aid output. Clinicians, therefore, can set a "wind noise program" with lower maximum output level and/or higher compression ratio for louder sounds to be used in environments where the hearing aid user does not need to converse much but may be bothered by wind noise (e.g., during jogging).

Besides the clinical applications, the results of this study also have several engineering implications. As mentioned before, an input limiter for digital hearing aids can simply be the upper limit of the dynamic range of the digital chip. When the overall sound pressure level of the incoming flow noise exceeded this level, digital distortions and unwanted noises were generated. It is, therefore, beneficial to use digital chips with higher input dynamic ranges to avoid or defer the digital distortions to a higher wind noise level/velocity.

In addition, an effective wind noise minimizing algorithm can be designed to take advantage of the voice detection and the limiting algorithms. When speech is not present, the thresholds of the output limiting algorithm can be lowered to reduce wind noise and increase comfort. When speech is present, the thresholds can be adjusted to allow speech clarity. For hearing aids with multichannel output limiting algorithms, the gain and/or the threshold of the output limiting algorithm for the low-frequency channel can also be lowered to reduce wind noise at low-frequency regions.

Further, although directional microphones generally yielded higher wind noise levels at the hearing aid output, engineering efforts can be applied to take advantage of their ability to reduce background noise in the far-field. An effective wind noise reduction algorithm can be implemented in conjunction with a multichannel automatic directional and omnidirectional microphone switching algorithm so that the hearing aid will always switch to the microphone mode with lesser output levels in each channel to reduce the overall noise level.

As wind noise is a rarely reported topic in literature, future studies are warranted in several areas: (1) the relationship among anatomic structures, flow characteristics, and flow noise in hearing aids; (2) the perceptual effects of wind noise on speech recognition and perceived sound quality for different hearing instrument users; (3) effects of signal processing algorithms; and (4) effect of physical modifications.

ACKNOWLEDGMENT

Many thanks are expressed to Paloma Mejia, Jong Beom Park, the Herrick Laboratories staff, and Dr. Kaibao Nie for their technical assistance. We also would like to thank Sonic Innovations for donating two behind-the-ear hearing aids for the study, and Michael Nilsson and Jennifer Grove for providing technical details on the hearing aids.

Anderson, J. D. (**2005**). *Fundamentals of Aerodynamics*, 4th ed. (McGraw-Hill Science Engineering, New York, NY).

Beard, J., and Nepomuceno, H. (**2001**). "Wind noise levels for an ITE hear-

ing aid," Knowles Engineering Report No. 128, Revision A., Itasca, IL.
Brown, D. V., and Mongeau, L. M. (**1995**). "The design, construction, and validation of a small, low-speed, quiet wind tunnel with application to noise from the flow over a cavity," Internal Report No. 204, HL95–9, Herrick Laboratories, Purdue University, West Lafayette, IN.
Burkhard, M. D., and Sachs, R. M. (**1975**). "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**, 214–222.
Chung, K. (**2004**). "Challenges and recent developments in hearing aids. Part I: Speech understanding in noise, microphone technologies and noise reduction algorithms," Trends Amplif. **8**, 83–124.
Grenner, J., Abrahamsson, U., Jernberg, B., and Lindblad, S. (**2000**). "A comparison of wind noise in four hearing instruments," Scand. Audiol. **29**, 171–174.
Kochkin, S. (**2000**). "MarkeTrak V: Why my hearing aids are in the drawer: The consumers' perspective," Hear J. **33**, 35–41.
Kochkin, S. (**2005**). "MarkeTrak VII: Customer satisfaction with hearing instruments in the digital age," Hear J., **38**, 30–37.
Leeuw, A. R., and Dreschler, W. A. (**1991**). "Advantages of directional hearing aid microphones related to room acoustics," Audiology **30**, 330–344.
National Climate Data Center, "Minimum wind noise," http://www.ncdc.noaa.gov/oa/climate/online/ccd/maxwnd.txt (Last viewed June, 2006a).
National Climate Data Center, "Minimum wind noise," http://www.ncdc.noaa.gov/oa/climate/online/ccd/maxwnd.txt (Last viewed June, 2006b).
Ricketts, T. A., and Hornsby, B. W. (**2003**). "Distance and reverberation effects on directional benefit," Ear Hear. **24**, 472–484.
Thompson, S., and Dillon, H. (**2002**). "Wind noise in hearing aids," presented at the American Academy of Audiology Convention, Philadelphia, PA.
Yip, M. K., "Revealing the magic in everyday: The science of soccer," www.physics.hku.hk/~phys0607/lectures/chap05.html (Last viewed May, 2007).

Improved fundamental frequency coding in cochlear implant signal processing

Matthias Milczynski,[a] Jan Wouters, and Astrid van Wieringen
ExpORL, K.U.Leuven, O & N 2 - Herestraat 49 bus 721, B-3000 Leuven, Belgium

(Received 6 August 2008; revised 28 January 2009; accepted 28 January 2009)

A new signal processing algorithm for improved pitch perception in cochlear implants is proposed. The algorithm realizes fundamental frequency (F0) coding by explicitly modulating the amplitude of the electrical stimulus. The proposed processing scheme is compared with the standard advanced combination encoder strategy in psychophysical music perception related tasks. Possible filter-bank and loudness cues between the strategies under study were minimized to predominantly focus on differences in temporal processing. The results demonstrate significant benefits provided by the new coding strategy for pitch ranking, melodic contour identification, and familiar melody identification.
 [DOI: 10.1121/1.3085642]

PACS number(s): 43.66.Ts, 43.66.Hg, 43.75.Cd [JCM] Pages: 2260–2271

I. INTRODUCTION

Pitch is an important perceptual property that is crucial for perception of musical melodies, gender discrimination in voice, prosody, and intonation of speech. In many cases, the fundamental frequency (F0) of a sound can be regarded as the physical counterpart of pitch. It is well known that cochlear implant (CI) recipients perform very poorly in pitch-related tasks and that this hampers music perception and appraisal. In Gfeller *et al.*, 2002 a familiar melody recognition test yielded significantly worse scores from CIs compared to normal hearing (NH) persons for both rhythmic and isochronous items. In the same study an average just noticeable difference in F0 (F0JND) of 7 semitones for a group of 46 CI users was determined. A control group of eight NH participants obtained considerably lower values (between one and two semitones). The high discrepancy in performance between CI and NH subjects was confirmed in other studies (Looi *et al.*, 2008; Fujita and Ito, 1999; Kong *et al.*, 2004). However, Nimmons *et al.* (2008) demonstrated that some CI subjects can achieve even smaller F0JNDs below one semitone.

In electrical hearing (EH) pitch can be mediated via temporal and place cues. Place cues refer to the place of stimulation along the basilar membrane. Expressed in terms of tonotopical organization, shifts from apex to base elicit a rise in pitch and vice versa. Temporal pitch can be transmitted either through variations of pulse rate (rate pitch) or envelope fluctuations of the stimulation pattern (envelope pitch), e.g., sinusoidal amplitude modulation (SAM) of current level at a fixed stimulation rate. In this case increasing the stimulation rate or modulation frequency, respectively, increases the perceived pitch. Furthermore, in the case of SAM pulse trains, an increase in modulation depth reinforces the periodicity information encoded by the modulation frequency and thus enhances the perceived pitch.

A large corpus of literature exists in which the potential of temporal pitch paradigms (and their possible interdependencies) was studied (Shannon, 1983; Townshend *et al.*, 1987; Pijl and Schwarz, 1995). McKay *et al.* (1994) showed that the pitch of SAM pulse trains (for carriers above 800 pps) can be matched with the pitch of unmodulated pulse trains presented at a stimulation rate similar to the modulation frequency. In other words, rate and envelope pitch can lead to a similar pitch percept. Both temporal pitch cues have a perceptual limit of about 300 Hz even though stimulation rate can elicit pitch percepts up to 1 kHz for some individuals (Shannon, 1983; Townshend *et al.*, 1987). At higher stimulation rates it is assumed that refractory properties of auditory nerve fibers become salient leading to an inconsistent pitch percept.

Nowadays, the commercially available processing strategies are based on the continuous interleaved sampling (CIS) (Wilson *et al.*, 1991) and advanced combination encoder (ACE) (Vandali *et al.*, 2000) processing schemes. For these algorithms an incoming signal is divided into a predefined number of frequency bands. This is accomplished by means of a filter bank with center frequencies usually spaced linearly below and logarithmically above 1 kHz. The envelope of each filter-bank channel is extracted either via half-wave rectification and lowpass filtering (CIS) or through magnitude response calculation and channel combinations (ACE). CIS and ACE differ mainly in the selection of valuable channels. In the CIS strategy usually all available electrodes are stimulated. In ACE, however, only a specified number of channels with maximal magnitudes are selected in every stimulation cycle. In subsequent processing steps the envelopes are subject to a nonlinear loudness growth function (LGF) and are finally used to modulate a fixed-rate electrical pulse train. Thus, the two pitch paradigms used in current processing strategies are place and envelope pitch. Processing schemes embedding variable stimulation rates such as F0F1F2 or MPEAK [see Loizou, 1998 for an excellent review] were predominantly represented in earlier commercial systems.

Several new experimental CI strategies targeting at improved F0 and/or music coding have been proposed in the present decade. Geurts and Wouters (2001) investigated the relevance of modulation depth in SAM pulse trains for modulation depth discrimination and F0JND of synthetic

[a]Electronic mail: matthias.milczynski@med.kuleuven.be

 0001-4966/2009/125(4)/2260/12/$25.00

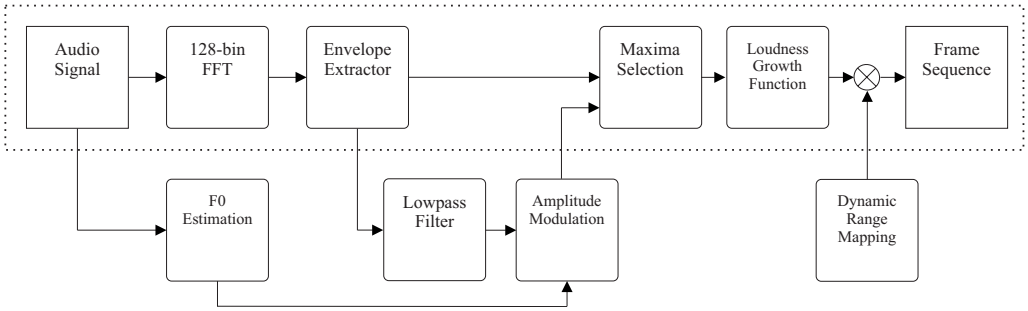


FIG. 1. Processing schemes of the standard ACE (components enclosed by dashed rectangle) and the modified F0mod (all components) strategy. The figure also depicts the location of the dynamic range (DR) mapping unit used in the loudness balancing procedure (see also Sec. III A).

vowels. A new processing strategy called F0CIS was implemented that allowed control over the modulation depth of temporal fluctuations. In the F0JND task no significant differences between F0CIS with increased modulation depth and the standard CIS scheme were demonstrated. Removal of temporal fluctuations (i.e., no available envelope pitch) resulted in the poorest performance.

Fearn (2001) developed a processing strategy with multiple channel stimulation rates. By employing an approach similar to the phase vocoder (Flanagan and Golden, 1966), instantaneous frequencies extracted from frequency bins of a digital fast fourier transform (FFT) filter bank were translated to stimulation rates in frequency channels below 1000 Hz. From the evaluation procedure no clear conclusion about the potential of the proposed strategy can be drawn (only two subjects were tested under different conditions).

A number of different experimental strategies were investigated by Vandali *et al.* (2005) and compared to the CIS and ACE strategies. In particular, the multichannel envelope modulation (MEM) strategy showed improvement in pitch ranking and no degradation in speech recognition (in quiet and in noise) as compared to ACE. Similar results for speech recognition were shown in a more recent study (Wong *et al.*, 2008). In their MEM approach envelope fluctuations, which inherently contain periodicity information, were extracted and additional control of modulation depth was provided. Modulation took place in phase across channels.

Green *et al.* (2004) proposed a coding strategy in which the filter-bank channel outputs of the CIS processing strategy were first lowpass filtered and then amplitude modulated at the F0 of the incoming sound with a sawtooth-shaped modulation. In a follow-up study (Green *et al.*, 2005) an improvement was demonstrated for a prosody identification task compared to the standard CIS strategy, but vowel recognition and formant frequency discrimination degraded with the new scheme. The authors concluded that the enhancement of temporal fluctuations might have an impact on spectral processing.

In a study by Laneau *et al.* (2006) the authors proposed a new coding strategy called F0mod that employed an enhanced coding of the envelope pitch cue. The strategy incorporated a F0 estimator into the processing framework of the ACE scheme. Filter-bank channels were amplitude modulated at the extracted F0. In a psychophysical evaluation significant advantages with F0mod over ACE were demonstrated in the scope of pitch ranking and melody recognition experiments.

In the present study we pursued the F0mod study, thereby focusing not only on implementation but also on psychophysical issues. Firstly, we aimed at a careful compensation of possible filter-bank cues that could have been an issue in the previous study due to the differences in spectral processing between ACE and F0mod. Secondly, in order to allow an even more systematic comparison between both schemes, possible differences in loudness were examined and eliminated in subsequent experimental procedures. Our hypothesis is that by minimizing spectral and loudness cues, differences between the strategies under study mainly result from their distinct temporal processing.

II. GENERAL METHODS

A. Modified F0mod strategy

We briefly review the most relevant components of the F0mod strategy as proposed in Laneau *et al.*, 2006, and the reader is kindly referred for further information not mentioned here. The strategy incorporated an autocorrelation-function-based F0 extractor into an ACE-like processing assembly. The channel outputs obtained from the filter bank were SAM at the estimated F0. The modulations were applied at full modulation depth (100% modulation) and in phase across channels. It was demonstrated in Laneau *et al.*, 2004 that these are optimal configurations for multichannel stimulation involving temporal cues. Moreover, the original F0mod algorithm from Laneau *et al.* (2006) introduced two modifications in spectral processing, as compared to ACE, namely, (1) a 512-bin FFT filter bank instead of a 128-FFT filter bank and (2) readjustment of the filter-bank center frequencies to allow for a better resolution in the low frequency range. The first modification was justified by the requirement that no temporal fluctuations in the filter-bank outputs should interfere with the amplitude modulations introduced by the F0 coding. The second modification was mainly motivated by outcomes from a study by Geurts and Wouters (2004) in which a filter bank with higher low frequency resolution could provide a better place pitch percept. In the evaluation procedure the original F0mod strategy was tested in comparison to standard ACE. CI subjects obtained statistically significant better scores with F0mod in pitch ranking of mu-

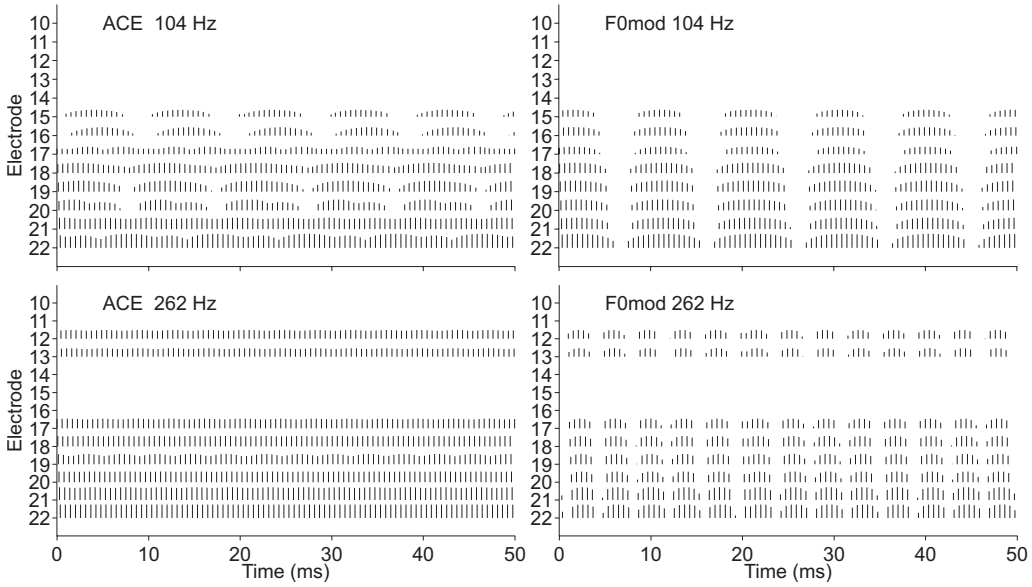


FIG. 2. Example electrodograms for harmonic complexes with F0's at G#2 (104 Hz) and C4 (262 Hz). The figures show stimulation patterns obtained with ACE and F0mod. Time is denoted on the *x*-axis of each figure and the electrode number (in base-to-apex order) on the *y*-axis. Strategy and F0 are given in the upper-left corner of each figure. Each electrodogram is truncated to 50 ms duration and to the range of active electrodes. The channel stimulation rate was 1800 pps in each example.

sical tones (for a limited frequency interval between 130 and 233 Hz) and in melody identification of isochronous items with F0mod. However, as explained above, not only enhanced temporal cues but also the modified spectral processing might have had an impact on the subjects' responses.

In order to minimize differences in spectral processing between ACE and F0mod a modified implementation is proposed here. In this implementation exactly the same filter bank as in the standard ACE scheme is used, i.e., a 128-bin FFT filter bank with the same crossover frequencies. The F0 estimation algorithm was enhanced by a voiced-unvoiced decision component. For voiced signals the envelopes of the filter-bank outputs were low pass filtered with a fourth order Butterworth filter (cutoff frequency at 60 Hz) and then amplitude modulated at the frequency of the extracted F0 (again at 100% modulation). Unvoiced sound segments were processed identically to ACE. The modified processing scheme is depicted in Fig. 1. If not otherwise stated the term F0mod will always refer to the new implementation.

In Fig. 2 stimulation patterns obtained from both strategies applied to two synthetic harmonic complexes (see also Sec. II C) at two different F0's are illustrated. Information regarding strategy and F0 (in hertz) is given in the upper-left corner of each electrodogram. The length of each segment was reduced to 50 ms and to the range of active electrodes for illustration purposes. The electrodograms clearly depict the similarities in peripheral information and the considerable differences in the temporal patterns between both schemes. In particular, temporal fluctuations that occur in ACE for the 104 Hz tone exhibit noticeable discrepancies in modulation depth, phase, and modulation shape. In contrast, the temporal information provided by F0mod is more regular with regard to these aspects. A striking difference in temporal processing between F0mod and ACE is also observed for the 262 Hz tone. In this case temporal information in ACE is almost completely absent, since this relatively high F0 exceeds the effective bandwidth of 250 Hz[1] of each FFT bin. In Fig. 3 average channel amplitudes are presented for the same two tones. The averaging was applied to each channel output of the LGF unit depicted in Fig. 1. Thereby, the amplitude values were mapped to the range [0, 1]. In general, larger average amplitudes are observed for ACE than for F0mod, in particular, for stimuli with higher F0's. This could result in possible loudness differences between ACE and F0mod, which will be addressed in Sec. III A.

B. Objectives of psychophysical tests

In the psychophysical evaluation we focused on music related tasks. The procedures were chosen under consideration of different aspects that are relevant to music perception. First of all, our evaluation test battery contained pitch ranking and familiar melody identification (FMI) of isochronous nursery songs. These two tests were also a part of the previous F0mod study and can be considered as standard procedures for the assessment of music perception in CIs (Nimmons *et al.*, 2008; Kasturi and Loizou, 2007; Kong *et al.*, 2004). These tests differ considerably in complexity with regard to perceptual dimensions. Pitch ranking demands a reasonable perception/comprehension of musical intervals and understanding of pitch. Indeed, it is not strictly related to perception of music since it can also be used for distinguishing between male and female voices. The FMI test bears a strong relation to music, but requires a certain degree of familiarity with the presented tunes. This is a critical issue since the ability to identify a musical piece is not only associated with perceptual abilities. The result will also depend on how well the particular item is represented internally (e.g., due to listening habits) and whether it can be connected to associative experiences (Peretz *et al.*, 2003). A further

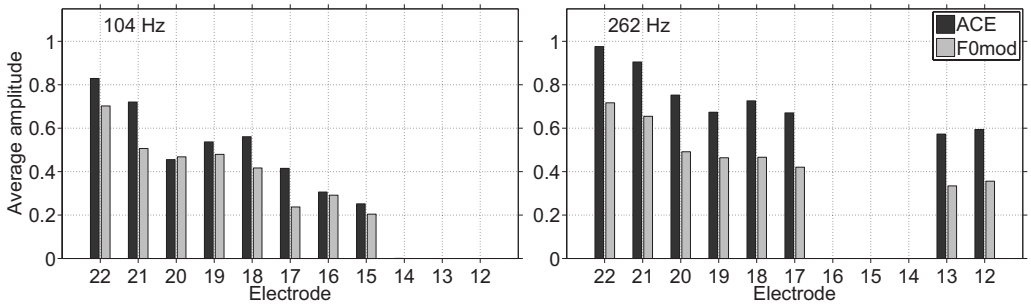


FIG. 3. Average amplitudes calculated from the electrodograms shown in Fig. 2. Note that the amplitude values at the output of the LGF are in the range between 0 and 1.

complication of the FMI test is that due to the modified rhythm the isochronous songs might become unfamiliar. In order to make a good compromise between both worlds we included the melodic contour identification (MCI) test (Galvin *et al.*, 2007) as part of the evaluation. The test addresses the perception of complex and music related stimuli but does not assume any familiarity with the stimuli. Thus purely musical abilities are accessed with the task. With these three tests (pitch ranking, MCI, and FMI) a composition of procedures is obtained with focus on different aspects of music perception at different levels of complexity.

C. Stimuli

The design of the stimuli in all procedures was identical to the synthetic harmonic complexes of the UW-CAMP test battery (Nimmons *et al.*, 2008). We used similar amplitude and phase information in the synthesis of the harmonic sounds. However, we did not apply the characteristic linear decay as was done in the UW-CAMP battery. The decay might influence the detection of envelope fluctuations perceptually since it is known that temporal modulation transfer functions (TMTFs) are level dependent in NH and EH (Shannon, 1992). For this reason we applied a 25 ms cosine ramp to the beginning and end of each sound with a 450 ms sustained part in between. The total duration of each stimulus was thus 500 ms. In this manner, a total number of 21 tones were generated with F0's ranging from 104 to 330 Hz, which is the range of most effective temporal pitch cues in CIs. Each complex stimulus contained 16 harmonics, meaning that the bandwidth of the stimuli ranged from 1500 (lowest F0 stimulus) to almost 5000 Hz (highest F0 stimulus).

The stimulus waveforms were all preprocessed in the nucleus MATLAB toolbox (NMT) developed by Cochlear Corp. This toolbox facilitates the conversion of digital waveforms into electrical stimulation patterns (i.e., electrodograms). The patterns are stored in a specific data structure that contains information about the active electrode, current level, stimulation mode, pulse width, phase gap, and time instance of stimulation for each cycle. The resulting data structure can be further used by particular software and hardware interfaces in order to be provided directly to a subject's implant (see also Sec. II D for details). The F0mod strategy was implemented in the framework of the NMT in which the standard ACE strategy is also available. Each electrical stimulus contained a channel stimulation rate of 1800 pps. This relatively high stimulation rate was necessary to allow a reasonable sampling of modulation frequencies above 250 Hz in F0mod (see also Wilson *et al.*, 1997). In each stimulation the maximal number of active channels was 8. The electrical pulses were biphasic with standard pulse width and phase gap values of 25 and 8 μs, respectively.

TABLE I. Details about CI subjects who participated in this study.

Subject	Age (yr)	Etiology	Duration of profound deafness (yr)	CI experience (yr)
S1	63	Progressive	>20	5
S2	60	Progressive	>30	7
S3	51	Unknown	6	12
S4	23	Meningitis	<1	5
S5	40	Progressive	<2	13

D. Setup

For all experimental procedures the APEX3 software (Francart *et al.*, 2008) served as the main platform for control, presentation of stimuli, and recording of the subjects' responses. The APEX3 framework also embeds the nucleus implant communicator (NIC) interface provided by Cochlear Corp. The interface facilitates direct streaming of the preprocessed electrodograms to a subject's implant. In doing so, the stimulation patterns, as described in Sec. III C, are sent directly to the internal part of the CI through a special clinical research processor. During each test session the subjects were presented with a graphical user interface. The possible answers for a particular task were represented by buttons, which had to be chosen by operating a mouse interface.

E. Subjects

Five postlingually deafened CI users participated in this study. Three subjects used the Nucleus Freedom speech processor and two subjects used the 3G speech processor. The strategy used by all subjects was ACE operating at 900 pps channel stimulation rate. The number of subjects varied per test. Subject S5 was not able to do the MCI test since he had a hard time understanding the task. Only three subjects (S1, S2, and S3) were familiar with the melodies used in the FMI test, so only those three subjects participated in that test. Table I provides details on the participating subjects.

III. EXPERIMENT 1: LOUDNESS BALANCING

A. Procedure

A loudness balancing procedure was conducted in order to equalize the loudness percept of all stimuli that were used in subsequent experiments. An adaptive 1-up, 1-down, 2-interval, 2-alternative-forced-choice (2I-2AFC) procedure was applied. The adaptive parameter of the procedure was the percentage of dynamic range (DR) to which the magnitudes of each channel were mapped. The parameter will be referred to as the DR mapping level (DRML). In this procedure, the DR is defined as the difference between most comfortable and threshold levels (or *T*-and *C*-levels) expressed in clinical current units. Mathematically, the DR mapping process for a channel n can be formulated as follows:

$$y_n = t_n + x_n(m/100)(c_n - t_n). \quad (1)$$

Thereby, x_n is the output of the LGF unit and the *T*- and *C*-levels are given by t_n and c_n. The DRML is represented by m that determines the percentage of DR to which the LGF output is mapped and which can obtain values between 0 and 100. Note that the DRML determines the peak amplitude of y_n. Figure 1 illustrates the DR mapping unit in the processing chain. In the test procedure, the subjects were presented with a fixed reference stimulus and a test stimulus in each trial in random order. They had to indicate the louder stimulus. The stimulus at 131 Hz F0 processed by ACE at a fixed DRML of 75% (thus $m=75$) served as the reference. In case the test stimulus was perceived as being louder its DRML was decreased; otherwise it was increased by a given amount Δm. The size of Δm was decreased from 10 to 2 throughout the experiment. After six reversals the procedure was terminated and the mean DRML was calculated from the last three reversals. Each test stimulus was played twice: once starting at $m=90$ and once at $m=60$. The mean of the two recorded DRMLs was taken as the final level at which the test stimulus was loudness balanced with the reference. Since the procedure is very time consuming the procedure was conducted for 11 tones ranging from 104 to 330 Hz, in steps of 1 whole tone, for both strategies. DRML values for intermediate tones were linearly interpolated.

The individual DRMLs (i.e., per tone and per subject) were used in all following experiments. Consequently, equal loudness was not only approached between strategies but also between stimuli within a strategy. Moreover, all tones were roved in loudness by adding a random value between −5% and +5% to the final DRML. Thus, the reference and test tokens in the pitch ranking test (see Sec. IV A) as well as contours and melodies in the MCI and FMI experiments (see Secs. V A and VI A), respectively, exhibited a different roving pattern in each trial. A similar loudness roving approach was applied by Nimmons *et al.* (2008) to acoustical stimuli. In the experiments described below none of the participating subjects reported apparent loudness effects.

B. Results

The results of the loudness balancing experiment averaged over five subjects are summarized in Fig. 4. Each marker represents the final DRML (y-axis) of a particular comparison stimulus for ACE and F0mod. The corresponding F0's (in hertz) are denoted on the x-axis. The dashed horizontal line highlights the DRML of the reference tone. There is a noticeable between-strategy difference in the progression of the loudness balance curves. In ACE DRMLs were adjusted to slightly higher values than the reference for F0's below 147 Hz but to progressively smaller values at higher F0's. A different trend is seen for F0mod. For F0's up to 165 Hz the loudness balance curve shows a steep positive slope starting at relative low DRMLs of about 55% and then approaching the DRML of the reference. At higher F0's the DRMLs were adjusted to values of about 80%. These results indicate that the loudness evoked by ACE stimuli increased with increasing F0 while the loudness of stimuli provided by F0mod decreased with increasing F0. A direct comparison of both strategies implies that at low F0's the F0mod stimuli are noticeably louder than in ACE but softer at higher F0's.

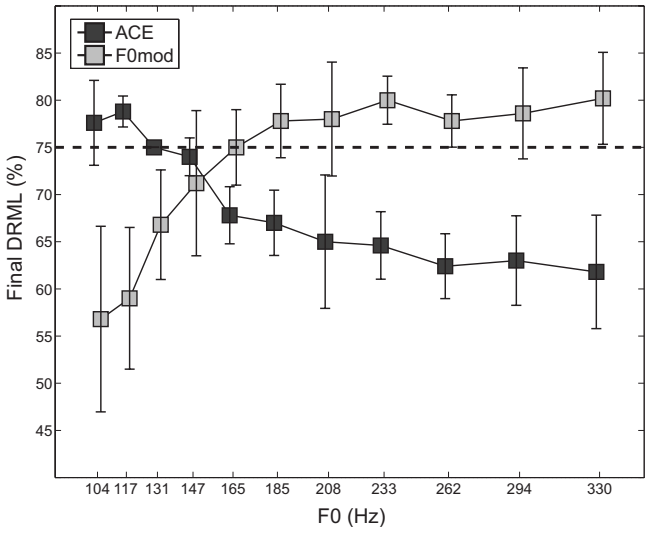


FIG. 4. Results loudness balancing. DRMLs (y-axis, see text for explanation of this parameter) obtained from the loudness balancing experiment averaged over five subjects are shown for ACE (dark gray markers) and F0mod (light gray squares). The particular tones are shown on the x-axis together with their corresponding F0's in hertz. The error bars indicate standard deviation and the dashed horizontal line highlights the DRML of the reference.

We performed a repeated measures ANOVA with two within-subject factors (strategy and F0) with the lower-bound adjustment for the degrees of freedom, on the final DRML values. The analysis revealed a significant effect of strategy ($p<0.05$). Moreover, the interaction between strategy and F0 was significant ($p<0.05$). The analysis also yielded a significant between-subject effect ($p<0.000\,001$).

C. Discussion

The results of experiment 1 demonstrated a significant difference in evoked loudness between both strategies under study. In both schemes the perceived loudness was frequency dependent. Since the spectral processing of ACE and F0mod was similar only differences in temporal processing could contribute to these outcomes. It is known from literature that temporal aspects influence the perceived loudness in EH. In Zeng and Shannon, 1995 the loudness of an AM stimulus

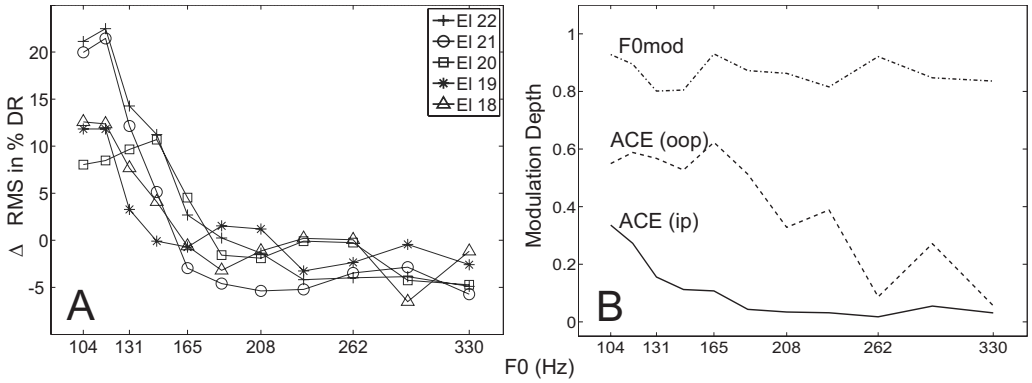


FIG. 5. (A) Differences in channel-rms values between ACE and F0mod expressed in % DR averaged over five subjects. The rms values were calculated from channel amplitudes (in microamperes) of electrodograms that yielded equal loudness in experiment III. Only those electrodes that were activated for each test stimulus were taken into account. (B) Modulation depth calculated from temporal information available in each electrodogram. For ACE the calculations were performed with [denoted by ACE (ip)] and without phase [dashed line denoted by ACE (oop)] information.

(modulation frequency was 100 Hz) was matched with a sinusoidal standard at 1 kHz in four Ineraid users. It was demonstrated that the slopes of the loudness balance functions decreased with increasing modulation depth indicating an increase in perceived loudness compared to an unmodulated stimulus. Moreover, the loudness difference between modulated and unmodulated signals was largest at high sensation levels. The authors concluded that either signal rms (low sensation levels) or signal peak (high sensation levels) determine the loudness percept. This was confirmed in a later study by Zhang and Zeng (1997) where Schroeder-phase harmonic complexes were balanced in loudness with zero-phase stimuli in three Ineraid users. In the same study an effect of modulation depth and modulation frequency on the loudness percept of SAM noise was shown in four NH subjects for relative low overall signal rms levels of up to 45 dB sound pressure level. Furthermore, it was reported by several authors (Cazals *et al.*, 1994; Chatterjee and Robert, 2001; Shannon, 1992) that TMTFs in EH are level dependent and dramatically decrease in slope (i.e., decrease in sensitivity) for modulation frequencies around 100 Hz and higher. These findings are relevant for our study. However, several differences that make a direct comparison somewhat more complicated should be pointed out. Firstly, in the above cited studies, only single electrodes were directly stimulated. In our study we use multichannel stimulation based on an entire implementation of a CI signal processing strategy. This means that we deal with by far more complex temporal and place characteristics of the electric stimulation pattern that vary not only across strategies but also across channels within a strategy. Secondly, the AM as implemented in our study was based on the following equation:

$$s(t) = \left[0.5 + 0.5 \sin\left(\frac{2\pi}{R}\sum_{i=0}^{t} F0(i)\right)\right] C_n(t),$$
$$n = 1 \cdots 22, \tag{2}$$

where $C_n(t)$ is the magnitude of filter-bank channel n at time t, $F0(i)$ is the extracted F0 at time i, R is the channel analysis rate (1778 Hz), and $s(t)$ is the resulting AM waveform. It is clear from Eq. (2) that the AM does not change the peak level of $C_n(t)$. Thirdly, we used pulsatile electric stimulation as opposed to analog stimuli used in Zeng and Shannon, 1995 and Zhang and Zeng, 1997. We suggest that this is the least critical issue concerning a comparison across studies since pulsatile and analog stimulations differ considerably in their physical properties but seem to evoke similar loudness percepts (see Zeng and Shannon, 1994).

We were interested in finding physical properties of the electric stimuli that would allow a reasonable interpretation of the perceptual results. Therefore, for each subject, we calculated the rms values in each channel of electrodograms that resulted from the loudness balance procedure. In each electrodogram we only considered those electrodes that were activated throughout the entire frequency range (i.e., in this case, electrodes 22–18). The rms values are first calculated from the channel amplitudes in μA and then expressed in % DR to allow a unified measure across subjects. In this case, the DR was calculated for each channel from its corresponding T- and C-levels in microamperes. The differences in channel-rms values between ACE and F0mod are shown in Fig. 5(A) for each of the selected electrodes. At low F0 stimuli (i.e., F0's in range 104–131 Hz) the channel-rms values of the selected electrodes are considerably higher in ACE than in F0mod. This is particularly true for the two lowest frequency electrodes 22 and 21. However, for F0's above 165 Hz the difference in channel-rms between both strategies decreased and was within 5%. The bias toward negative differences can be explained by the nonlinear relationship between microamperes and clinical current units that were used to express DR in the loudness balancing experiment. In other words, for this F0 region, the channel-rms values seem to be related to loudness. However, this measure cannot explain the high loudness discrepancy below 165 Hz. We speculate that in this case across-channel phase synchrony in AM and the subjects' frequency dependent modulation sensitivity have an effect on loudness.

Panel (B) of Fig. 5 shows modulation depths (range [0, 1]) calculated from the spectra of the channel amplitudes of each electrodogram [see Eq. (5) in Laneau *et al.*, 2004]. For ACE we also made a distinction between in-phase and out-of-phase modulations. From these measures it is obvious that ACE exhibits in-phase modulations in a limited frequency range at noticeably lower modulation depths than F0mod. It

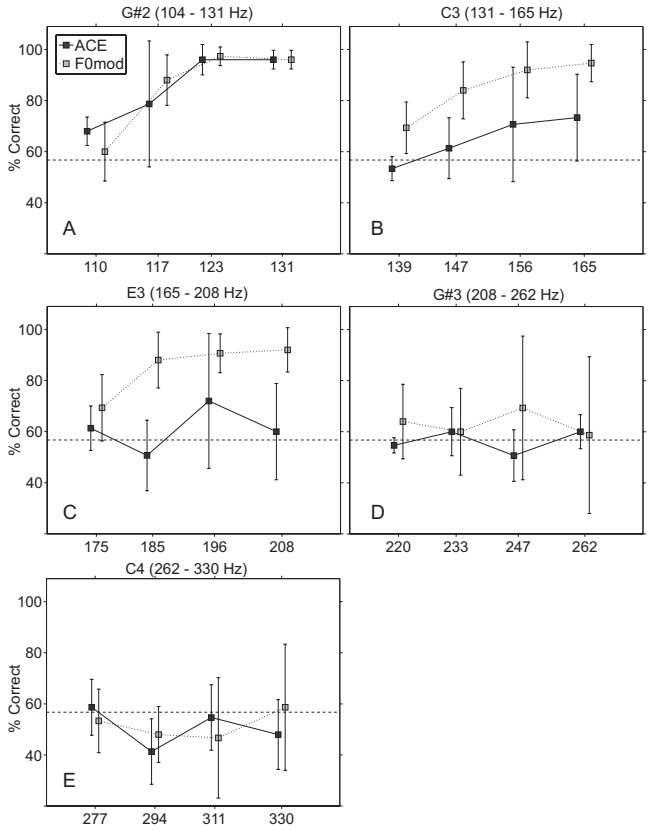


FIG. 6. Results pitch ranking. Panels (A)–(E) show the % correct scores (*y*-axis) for a particular frequency register (denoted by the title above each panel) averaged over five subjects. The F0's of the comparison tones (in hertz) are indicated on the *x*-axis. The reference note of the frequency register and the covered frequency range in hertz are denoted above each graph. The horizontal dashed line represents the level at which scores were significantly above chance level (with $p=0.05$ and chance level at 50%).

appears unlikely that in ACE these in-phase modulations could significantly contribute to a loudness percept. In contrast, F0mod contains in-phase modulations at full depth for the entire investigated frequency range. We assume that the only significant influence of these modulations was on loudness percept for F0's between 104 and 165 Hz. In summary, we propose that at low F0 stimuli mainly across-channel synchronized temporal fluctuations in the electric code contributed to the loudness percept for F0's between 104 and 165 Hz. At higher F0's the channel-rms levels essentially determined loudness. The implications of out-of-phase modulations on loudness remain unclear.

The demonstrated effects of the F0mod strategy on loudness could have a negative impact on speech perception. The different loudness percepts between ACE and F0mod could produce different consonant-vowel intensity ratios in both strategies. In F0mod, this might lead to less consonant enhancement (in case vowel and voiced consonants would appear equally loud) and consequently to degradation in consonant recognition (see also Kennedy *et al.*, 1998). We furthermore speculate that the evoked loudness is independent of modulating waveform since detection thresholds for, e.g., sinusoidally and sawtooth-shaped modulated pulse trains are similar as shown by Landsberger (2008). However, in order to overcome an eventual speech perception, deficit, subjects should be given sufficient time to get accustomed to the new processing scheme.

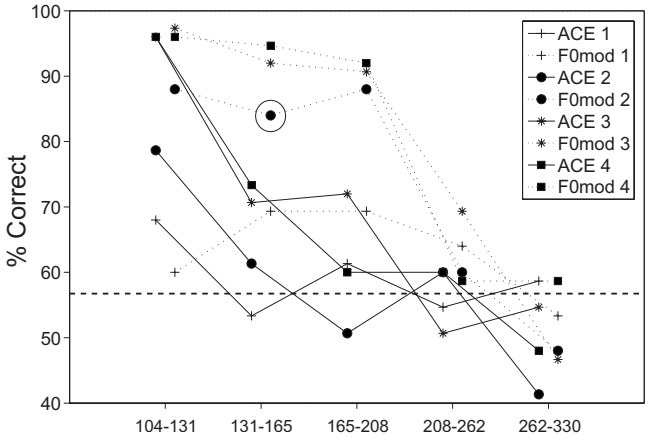


FIG. 7. Semitone-interval scores in which the pitch ranking results (experiment 2) were reorganized. The figure shows the % correct scores (*y*-axis) for particular intervals spanned by reference and comparison tones of each frequency register. Frequency registers are represented by their corresponding frequency range (in hertz) on the *x*-axis (compare with titles above each graph of Fig. 6). The parameters of this representation are strategy and interval (in semitones) and are denoted in the legend in the same order. Error bars were omitted for clarity. The circle exemplifies the % correct F0mod score for a two-semitone-interval spanned by reference and comparison tones in frequency range 131–165 Hz corresponding to frequency register C3. Note that semitone-interval scores representing the same interval and strategy are connected by a line.

IV. EXPERIMENT 2: PITCH RANKING

A. Procedure

Pitch ranking was conducted in a constant 2I-2AFC procedure. Upon each trial subjects were presented with a reference and a comparison tone in random order and had to indicate which was higher in pitch. A total of five frequency registers were created with reference F0's at 104, 131, 165, 208, and 262 Hz, respectively. Each register contained four comparison tones that were one, two, three, or four semitones higher in F0 than the reference. All frequency registers were tested separately in one block. Within a block each combination of reference and comparison tones was repeated 15 times resulting in a total of 60 trials (4 comparison tones $\times$ 15 repetitions) per block and thus per frequency register. No feedback on the subjects' responses was provided; likewise neither training nor practice sessions were presented.

B. Results

The pitch ranking results are shown in Fig. 6. Panels (A)–(E) depict the % correct scores for one particular frequency register represented by the corresponding reference note and frequency range (denoted above each graph). One filled square represents results averaged over responses from five subjects for one comparison tone (comparison tones are represented on the *x*-axis by their corresponding F0 in hertz). The dashed horizontal line in each graph shows the level at which a score significantly above chance is achieved (for $p=0.05$ and chance level at 50%). In Fig. 7 the pitch ranking results were reorganized into so called semitone-interval scores. Each marker shows the % correct score for a defined

interval spanned by reference and comparison stimuli from a given frequency register. Frequency registers are represented by their corresponding frequency range (in hertz) on the *x*-axis (compare with titles above each graph of Fig. 6). An example is given by the circle that represents the F0mod score for a two-semitone interval (i.e., the F0's were 131 and 147 Hz) in the frequency range 131–165 Hz corresponding to frequency register C3. In the figure legend the number following the strategy name denotes the interval in semitones. Note that each particular semitone-interval score (i.e., representing one interval and strategy) is connected by a line. This representation is particularly useful for an interpretation of the results of experiment 3 and will be referred to in Sec. V C.

A repeated measures ANOVA was performed on the scores taken from all registers with three within-subject factors (strategy, register, and comparison note). The analysis yielded significant effects for all factors ($p<0.05$). No other significant interaction effects were indicated. A pairwise comparison using the Bonferroni test showed that performance with F0mod was significantly better than with ACE ($p<0.05$). There was a significant between-subject effect ($p<0.0001$). A repeated measures ANOVA with two within-subject factors (strategy and comparison note) was applied to each frequency register separately. The analysis yielded a significant effect of strategy for registers C3 and E3. A pairwise comparison showed that indeed subjects performed significantly better with F0mod than with ACE ($p<0.01$) for these two F0 regions. The *p*-scores of the remaining results were not significant for both strategies.

C. Discussion

The results of experiment 2 demonstrated (1) equal performance with both strategies in register G#2, (2) a significantly better performance with F0mod in registers C3 and E3, (3) near chance level performance with both strategies in register G#3 with a small advantage for F0mod (average scores lie all significantly above chance level), and (4) overall low performance (i.e., not significantly above chance level in most instances) with both strategies in the last frequency register C4.

These outcomes can be best interpreted on the basis of three aspects that are relevant for perception of temporal pitch cues in EH, namely, (1) sensitivity to AM, (2) across-channel synchrony and depth of AMs, and (3) perceptual limit of temporal pitch cues. In frequency register G#2 the performance is equal in both strategies. In this frequency range, however, ACE exhibits in-phase modulations at considerably lower depths than in F0mod [see Fig. 5(B)]. This implies that at low F0's, where modulation sensitivity is highest, less modulation depth in AM synchronized channels than applied in F0mod is sufficient for pitch ranking. In frequency register C3, however, the scores with ACE drop compared to register G#2. This is probably due to the reduction in in-phase modulation depth for this strategy observed between 104 and 165 Hz. The out-of-phase modulation depth is almost constant in the same frequency range. The outcomes for ACE in register E3 can be explained in a similar way. Note that the performance with ACE in register C3 is more consistent (improved performance with increasing comparison F0) than in register E3. In contrast, the phase-synchronized AMs as provided by F0mod allow reliable pitch ranking between 104 and 208 Hz.

In the last two registers G#3 and C4, where the subjects exhibit least sensitivity to AM, temporal envelope cues become less important for pitch perception. In addition, in these registers, the generally accepted limit of temporal pitch cues, namely, 300 Hz, is approached. This means that in this case the subjects mainly relied on place pitch that probably could not contribute to a pitch percept alone within the limited frequency range of the above mentioned registers. However, in register G#3, the average scores with F0mod lie significantly above chance level showing that some subjects could indeed still use the temporal information to manage the task. This was confirmed by the individual scores of subjects S3–S5 who could clearly profit from F0mod compared to ACE especially for comparison F0's at 247 and 262 Hz (data not shown here).

Based on our results and previously reported findings (see also Sec. III C) we assume that with increasing F0 the subjects showed decreasing sensitivity to AM. In the F0mod strategy this diminishment could be compensated for F0 values up to 208 Hz by enhanced temporal cues, realized through across-channel synchrony in AMs applied at 100% modulation depth. The overall progression (between registers) of the pitch ranking scores together with modulation depth measures shown in Fig. 5 support the assumption that synchrony may be a more relevant property than modulation depth.

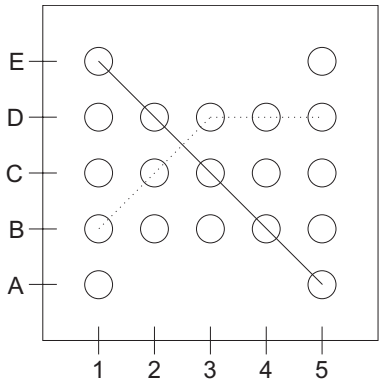
In Vandali *et al.*, 2005 several strategies that incorporate synchronized and enhanced envelope cues in a similar way to F0mod were compared with ACE. In a frequency range between 98 and 277 Hz, thus comparable to our study, significantly improved pitch ranking was demonstrated with the MEM, MDE2, and MDE3 strategies. The authors speculated that across-channel AM synchrony contributes more to an improved pitch percept than modulation depth. This in general agrees with our results. Furthermore, the authors also reported decreasing performance with increasing F0 another finding that is consistent with our study. It should be mentioned, however, that in the pitch ranking tests of the study by Vandali *et al.* (2005) the smallest interval between reference and comparison tones was half an octave. Thus, it remains unrevealed how well CI subjects would perform with their proposed strategies for smaller intervals that were assessed in our study.

V. EXPERIMENT 3: MELODIC CONTOUR IDENTIFICATION (MCI)

A. Procedure

The MCI test contained nine different melodic sequences of which each consisted of five tones. Every sequence was characterized by a specific F0 contour. The test was carried out in a constant 1I-9AFC procedure. In one trial the subjects were presented with one of the nine contours and were asked to indicate a corresponding visual counter-

Grp4	Grp3	Grp2	Grp1
262	208	165	131
208	175	147	123
165	147	131	117
131	123	117	110
104	104	104	104



Contour	Sequence
R	1A, 2B, 3C, 4D, 5E
FA	1E, 2D, 3C, 4B, 5A
FL	1C, 2C, 3C, 4C, 5C
RFA	1B, 2C, 3D, 4C, 5B
RFL	1B, 2C, 3D, 4D, 5D
FAFL	1D, 2C,3B, 4B, 5B
FAR	1D, 2C, 3B, 4C, 5D
FLR	1B, 2B, 3B, 4C, 5D
FLFA	1D, 2D, 3D, 4C, 5B

FIG. 8. Illustration of contours used in the MCI test. The contour matrix in the middle of the figure allows to navigate through the progression of each contour in time (horizontal locations 1, 2, 3, 4, and 5) and F0 (vertical locations A, B, C, D, and E). The table on the left contains the corresponding F0's (in hertz) of each contour tone for each interval group (denoted by Grp1–Grp4). These data are vertically aligned with the F0 locations in the contour matrix. For example, the point 4B corresponds to a F0 of 117 Hz in interval group 2 and a F0 of 131 Hz in interval group 4. The table on the right provides the sequence of contour matrix locations for each melodic contour. The following abbreviations were used: *R* (rising), FA (falling), FL (flat), RFA (rising-falling), RFL (rising-flat), FAR (falling-rising), FAFL (falling-flat), FLR (flat-rising), and FLFA (flat-falling). The solid and dashed lines in the contour matrix represent the FA and RFL contours, respectively.

part (Galvin *et al.*, 2007). We implemented four interval groups in which an interval of one, two, three or four semitones, respectively, separated two successive notes of one melodic contour. Each interval group was tested separately in one block, in which the order of contours was randomly chosen and each contour was repeated twice. Each block was repeated three times, yielding a total number of six presentations of one contour in one interval group. This resulted in 54 trials (9 contours $\times$ 6 repetitions) per interval group. Again, no feedback or training was provided. The particular tone sequences of each melodic contour and of each register can be derived using Fig. 8. The contour matrix in the middle of the figure allows to navigate through each contour in time (x-axis) and F0 (y-axis). The corresponding tones per register are given in the table on the left whose entries are aligned vertically with the y-axis of the contour matrix. The table on the right shows the sequence of tones for each contour as matrix locations. Two examples, FA and RFL, are depicted in the contour matrix (see caption of Fig. 8 for further explanation).

B. Results

In Fig. 9 the MCI test results representing averaged scores from subjects S1–S4 are shown. Panels (A)–(D) of Fig. 9 represent each interval group with bars indicating the % correct scores for one particular contour. The dashed line indicates the level at which scores were significantly above chance level (with $p=0.05$ and chance level at 11.1%). In Fig. 10 the scores for each interval group shown in Fig. 9 were averaged over the nine contours. In both figures the error bars indicate standard deviation.

A repeated measures ANOVA with three within subject factors (strategy, interval group, and contour) was performed on the scores taken from all interval groups. The analysis yielded a significant effect of strategy ($p<0.01$). The pairwise comparison showed a significantly better performance with F0mod. The between-subjects effect was significant ($p<0.01$). For each interval group a repeated measures ANOVA with two within-subject factors (strategy and contour) was performed. No statistically significant effect was indicated for any factor in interval groups 1 and 2. A statistically significant effect of strategy ($p<0.05$) was obtained in interval group 3 in which a pairwise comparison yielded a significantly better score obtained with F0mod. For the last interval group the better performance with F0mod approached significance ($p=0.055$).

We furthermore examined whether the subjects significantly improved their performance with one of the strategies across interval groups. We therefore performed a repeated measures ANOVA with two within-subject factors (interval group and contour) for each strategy. No significant effect for any of the factors was obtained for ACE. For F0mod the effect of interval group approached significance ($p=0.052$). A pairwise comparison yielded an improvement in performance from interval groups 1 to 4 that approached significance at the same p-level.

C. Discussion

In the MCI experiment we expected to observe an improved performance in higher interval groups as compared to lower ones since large intervals between successive tones require less pitch resolution for recognition. From Figs. 9 and 10 it is clear that performance with F0mod comes close to these expectations. The subjects consistently improved their scores with increasing interval group. Comparing interval groups 1–4 they, in fact, improved from 27% to 61% correct. In contrast, the progression of the scores obtained with ACE between interval groups is counterintuitive since the scores seem to decrease from the one-semitone (38% correct) to the three-semitone condition (20% correct). Though, in interval group 4 the performance with ACE was more accurate as compared to interval group 3 (improvement from 20% to 32% correct). Interestingly the results with ACE are insignificantly higher than with F0mod in the first interval group [see Figs. 10 and Fig. 9(a)].

We can carefully interpret the outcomes of the MCI task, in which recognition of intervals was crucial, using the semitone-interval scores shown in Fig. 7. The progressions

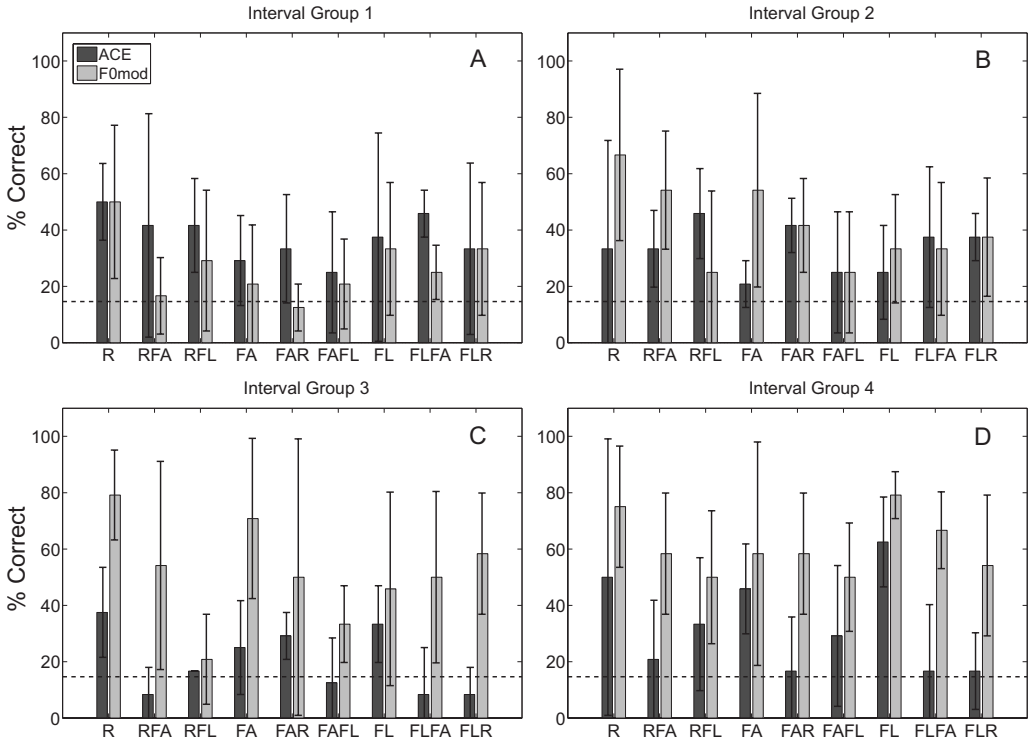


FIG. 9. Results MCI. The graphs in (a)–(d) for % correct scores (y-axis) for a particular interval group (denoted above each graph). Contours are denoted on the x-axis. The horizontal dashed line in each figure indicates the level of performance significantly above chance (for $p=0.05$ and chance level at 11.1%). Error bars indicate standard deviation.

of the semitone-interval curves of F0mod correspond with the average results of Fig. 10. First of all, an improvement is observed for this strategy between interval groups 1 and 2 (increase from 26.9% correct to 41.2% correct). The corresponding one- and two-semitone-interval scores also show an improvement by almost 30% (60% correct for one-semitone-interval score in frequency range 104–131 Hz, 88% and 84% correct for two-semitone-interval score in frequency range between 104 and 165 Hz). Moreover, an increase in performance is seen between interval groups 2 and 3 (improvement by 10%). This again agrees with the improvement between the two- and three-semitone-interval curves in the frequency range 104–208 Hz. Note also that these two curves show very little variance within this frequency range indicating that these two intervals could be perceived equally well throughout an entire melodic contour. Finally, the highest average scores with F0mod are obtained in interval group 4. The dramatic decrease in the corresponding four-semitone-interval scores from 208 Hz on had only little effect on the overall performance in this interval group since the only contours that spanned the entire frequency range between 104 and 262 Hz (thus from registers G#2 to G#3) were the R and FA contours. In fact, the average scores obtained for these contours are slightly lower than in interval group 3 [compare with panels (C) and (D) of Fig. 9].

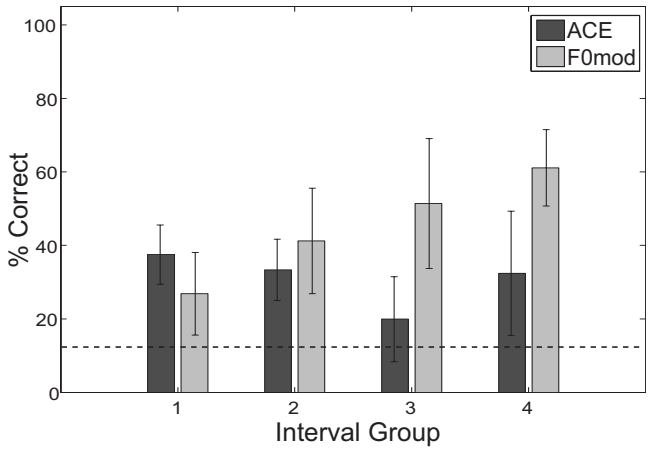


FIG. 10. Percent correct MCI scores (y-axis) averaged over four subjects and nine contours shown for each interval group (x-axis). Error bars indicate standard deviation.

The situation for ACE is somewhat different. The one-semitone-interval scores in frequency range 104–131 Hz for ACE amount to 69% correct. This could have led to the slightly better performance than F0mod in interval group 1. The two-semitone-interval scores for ACE, however, show a dramatic decrease from frequency range 104–131 Hz to 131–165 Hz (from 79% to 61%, respectively). This indicates that parts of the melodic contours falling between 104 and 131 Hz were more accurately recognized than parts falling between 131 and 165 Hz. An examination of the confusion matrices obtained with ACE for interval group 2 confirmed that this was indeed the case. The R contour, for instance, was eight times correctly identified but ten times confused with the RFL contour. Moreover the FA contour was 5 times correctly identified but 12 times confused with the FLFA contour. Similar confusions were observed in interval groups 3 and 4 for ACE. It remains, however, not entirely clear why the performance with ACE in interval group 3 was lower (even though not significantly) as compared to the remaining

registers. We speculate that up to interval group 3 temporal cues mainly mediated a pitch percept while for interval group 4 place cues prominently transmitted pitch information of higher F0's. As already discussed in Sec. IV C the temporal pitch cues in ACE systematically diminished from low to high F0's. For interval group 3 this would imply that as higher F0's were approached the subjects could neither reliably extract place nor temporal information with ACE.

In Galvin *et al.*, 2007 12 CI subjects were tested in the MCI task. In this study, among others, an interval group of five semitones was assessed. In many subjects consistently increasing scores with increasing interval group were observed. In contrast to this, in two subjects no considerable improvement was demonstrated between interval groups 2 and 5, one subject showed no improvement in performance between interval groups 2 and 4, and another subject could not improve noticeably between interval groups 3 and 5. Only one subject showed an inconsistent performance drop from interval group 3 to 4. However, in this test only frequency ranges above 220 Hz were tested. Furthermore, the study of Galvin *et al.* (2007) used stimuli containing only 3 harmonics, while our study used more complex stimuli containing energy in and up to the 16th harmonic. We assume that these are crucial differences to our study since harmonic complexes at higher F0's, with only a few harmonics, provide a distinct place pitch cue, which was also demonstrated by the study of Galvin *et al.* (2007).

VI. EXPERIMENT 4: FAMILIAR MELODY IDENTIFICATION (FMI)

A. Procedure

The FMI test contained two parts. In the first one the subjects were presented with 19 different titles of nursery songs well known to the Dutch–Flemish population. Thereby, they should indicate their familiarity with the title on a scale between 1 and 10 (1=not familiar, 10=familiar). The ten most familiar tunes were chosen. In the second part, one of the ten chosen familiar tunes was presented at random to the subjects in an isochronous version. Isochronous in this context means that rhythmical information was eliminated from the melodies by presenting each tone at equal duration and at equal time intervals. This approach prevents the subjects from using rhythm information as an additional cue. The subjects had to indicate the correct title of the song from a closed set of ten possible answers. Each melody was repeated twice within a block. One block was repeated three times yielding a total number of six presentations of each song. Again, as in the previous procedures, the subjects did not receive any feedback or training sessions. For more information about the song titles, the reader is referred to Laneau *et al.* 2006.

B. Results

The results of the FMI test are shown in Fig. 11 for each subject (S1–S3) and also for the group represented by the rightmost bars. Scores significantly above chance level are represented by the dashed horizontal line, and standard deviations are added to the group scores. A repeated measures ANOVA performed on the average scores did not indicate a significant effect of strategy. However, with F0mod, all subjects could obtain a score significantly above chance level, which was not the case for the ACE score of subject S3.

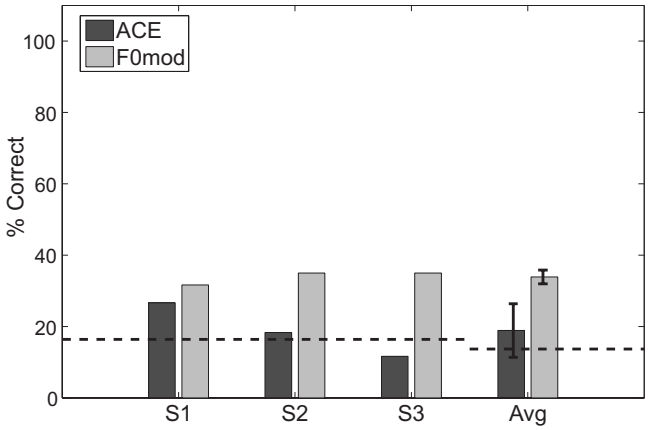


FIG. 11. Results FMI. The % correct scores (y-axis) are shown for each subject (denoted by S1–S3 on the x-axis). The two rightmost bars show scores averaged over the three participants. Note that the level at which performance was significantly above chance (dashed horizontal line, $p=0.05$ and chance level at 10%) is lower for the average scores.

C. Discussion

The results of the FMI task were modest. The subjects achieved scores between 31% and 35% with F0mod. The scores of two subjects (S2 and S3) were considerably smaller with ACE than with F0mod; in fact, subject S3 did not perform significantly above chance level with ACE. It is well known that in a FMI task with isochronous items CI users perform at chance level, which was shown, e.g., in Kong *et al.*, 2004 for four out of six CI subjects in a melody identification task. Overall low scores obtained by implanted subjects for recognition of isochronous melodies were also demonstrated in Gfeller *et al.*, 2002. This agrees with the results that were obtained for the ACE strategy in our study.

In general, in melody recognition tasks, CI subjects mainly rely on rhythmical cues rather than the melody itself. With F0mod the subjects could already obtain a baseline performance that was significantly above chance level. Note that the subjects did not have the opportunity to get accustomed to the sounds provided by F0mod. This means that with sufficient practice sessions eventually higher scores with F0mod could be demonstrated. This would imply that CI users would have better access to the melodic component of a familiar tune rather than to the rhythmic information alone. For the FMI tasks we should therefore gather data with more CI subjects over a longer term.

VII. CONCLUSIONS

In the present study we proposed an alternative implementation of a new coding strategy called F0mod originally developed in Laneau *et al.*, 2006. The strategy provides F0 information via amplitude fluctuations in each channel of the electrical stimulus thereby enhancing the envelope pitch cue. The new processing scheme was compared with ACE, nowadays the clinical standard strategy implemented in speech

processors of Cochlear Corporation. The evaluation was conducted in the scope of psychophysical tasks relevant to music perception. We carefully minimized possible differences in spectral processing as well as in evoked loudness percepts between the strategies under investigation. Thereby, we ensured that the strategies primarily differed in temporal processing. Significant improvements in pitch ranking, MCI, and FMI tests were demonstrated with the new F0mod scheme. The results indicate a potential benefit from consistently mediated envelope pitch cues, i.e., via amplitude modulations provided in phase across channels and at full modulation depth (i.e., 100% modulation). In particular, we also demonstrated that these cues significantly affect the loudness percept of a stimulus. This has to be taken into account in evaluation procedures and also when considering a real-time implementation.

ACKNOWLEDGMENTS

This research was carried out in the frame of the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT) Project No. 050445 and was partly sponsored by Cochlear Ltd. The authors thank the test subjects for their patient and enthusiastic participation to the experiments.

[1]This is the effective bandwidth of the hanning window, which is used in the FFT filter bank, for a length of 128 samples at a sampling rate of 16 kHz.

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (**1994**). "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," J. Acoust. Soc. Am. **94**, 2048–2054.

Chatterjee, M., and Robert, M. E. (**2001**). "Noise enhances modulation sensitivity in cochlear implant listeners: Stochastic resonance in a prosthetic sensory system?," J. Assoc. Res. Otolaryngol. **2**, 157–171.

Fearn, R. (**2001**). "Music and pitch perception of cochlear implant recipients," Ph.D. thesis, University of New South Wales, New South Wales, Australia.

Flanagan, J. L., and Golden, R. (**1966**). "Phase vocoder," Bell Syst. Tech. J. **45**, 1493–1509.

Francart, T., van Wieringen, A., and Wouters, J. (**2008**). "APEX 3: A multi-purpose test platform for auditory psychophysical experiments," J. Neurosci. Methods **172**, 283–293.

Fujita, S., and Ito, J. (**1999**). "Ability of nucleus cochlear implantees to recognize music," Ann. Otol. Rhinol. Laryngol. **108**, 634–640.

Galvin, J. J., Fu, Q. J., and Nogaki, G. (**2007**). "Melodic contour identification by cochlear implant listeners," Ear Hear. **28**, 302–319.

Geurts, L., and Wouters, J. (**2001**). "Coding of fundamental frequency in continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **109**, 713–726.

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

Gfeller, K., Fearn, R., Knutson, J., and Stordahl, J. (**2002**). "Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults," Coch. Imp. Int. **3**, 29–53.

Green, T., Faulkner, A., and Rosen, S. (**2004**). "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants," J. Acoust. Soc. Am. **116**, 2298–2310.

Green, T., Faulkner, A., Rosen, S., and Macherey, O. (**2005**). "Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification," J. Acoust. Soc. Am. **118**, 375–385.

Kasturi, K., and Loizou, P. C. (**2007**). "Effect of filter spacing on melody recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **122**, 29–34.

Kennedy, E., Levitt, H., Neuman, A. C., and Weiss, M. (**1998**). "Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners," J. Acoust. Soc. Am. **103**, 1098–1114.

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.

Landsberger, D. M. (**2008**). "Effects of modulation wave Shape on modulation frequency discrimination with electrical hearing," J. Acoust. Soc. Am. **124**, 21–27.

Laneau, J., Moonen, M., and Wouters, J. (**2004**). "Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees," J. Acoust. Soc. Am. **116**, 3606–3619.

Laneau, J., Moonen, M., and Wouters, J. (**2006**). "Improved music perception with explicit pitch coding in cochlear implants," Audiol. Neuro-Otol. **11**, 38–52.

Loizou, P. (**1998**). "Mimicking the human ear," IEEE Signal Process. Mag. **15**, 101–130.

Looi, V., McDermott, H. J., McKay, C. M., and Hickson, L. M. (**2008**). "Music perception of cochlear implant users compared to that of hearing aid users," Ear Hear. **25**, 421–434.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," J. Acoust. Soc. Am. **96**, 2664–2673.

Nimmons, G. L., Kang, R. S., Drennan, W. R., Longnion, J., Ruffin, C., Worman, T., Yueh, B., and Rubinstein, J. T. (**2008**). "Clinical assessment of music perception in cochlear implant listeners," Otol. Neurotol. **29**, 149–155.

Peretz, I., Champod, S., and Hyde, K. (**2003**). "Varieties of musical disorders: The montreal battery of evaluation of amusia," Ann. N.Y. Acad. Sci. **999**, 58–75.

Pijl, S., and Schwarz, D. W. F. (**1995**). "Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes," J. Acoust. Soc. Am. **98**, 886–895.

Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157–189.

Shannon, R. V. (**1992**). "Temporal modulation transfer functions in patients with cochlear implants," J. Acoust. Soc. Am. **91**, 2156–2164.

Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**). "Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–115.

Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W. D., and McDermott, H. J. (**2005**). "Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies," J. Acoust. Soc. Am. **117**, 3126–3138.

Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). "Speech perception as a function of electrical stimulation rate: Using the nucleus 24 cochlear implant system," Ear Hear. **21**, 608–624.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.

Wilson, B. S., Finley, C. C., Lawson, D. T., and Zerbi, M. (**1997**). "Temporal representations with cochlear implants," Am. J. Otol. **18**, 30–34.

Wong, L. L., Vandali, A. E., Ciocca, V., Luk, B., Ip, V. W., Murray, B., Yu, H. C., and Chung, I. (**2008**). "New cochlear implant coding strategy for tonal language speakers," Int. J. Audiol. **47**, 337–347.

Zeng, F.-G., and Shannon, R. V. (**1994**). "Loudness-coding mechanisms inferred from electric stimulation of the human auditory system," Science **264**, 564–566.

Zeng, F.-G., and Shannon, R. V. (**1995**). "Loudness of simple and complex stimuli in electric hearing," Ann. Otol. Rhinol. Laryngol. **166**, 235–238.

Zhang, C., and Zeng, F.-G. (**1997**). "Loudness of dynamic stimuli in acoustic and electric hearing," J. Acoust. Soc. Am. **102**, 2925–2934.

Characteristics of air puffs produced in English "pa": Experiments and simulations

Donald Derrick[a)]
Department of Linguistics, University of British Columbia, Totem Field Studios, 2613 West Mall, Vancouver, British Columbia V6T 1Z4, Canada

Peter Anderson
Department of Mechanical Engineering, University of British Columbia, 2054-6250 Applied Science Lane, Vancouver, British Columbia V6T 1Z4, Canada

Bryan Gick
Department of Linguistics, University of British Columbia, Totem Field Studios, 2613 West Mall, Vancouver, British Columbia V6T 1Z4, Canada and Haskins Laboratories, New Haven, Connecticut 06511

Sheldon Green
Department of Mechanical Engineering, University of British Columbia, 2054-6250 Applied Science Lane, Vancouver, British Columbia V6T 1Z4, Canada

(Received 31 March 2008; revised 20 January 2009; accepted 21 January 2009)

Three dimensional large eddy simulations, microphone "pop" measurements, and high-speed videos of the airflow and lip opening associated with the syllable "pa" are presented. In the simulations, the mouth is represented by a narrow static ellipse with a back pressure dropping to 1/10th of its initial value within 60 ms of the release. The simulations show a jet penetration rate that falls within range of the pressure front of microphone pop. The simulations and high-speed video experiments were within 20% agreement after 40 ms, with the video experiments showing a slower penetration rate than the simulations during the first 40 ms. Kinematic measurements indicate that rapid changes in lip geometry during the first 40 ms underlie this discrepancy. These findings will be useful for microphone manufacturers, sound engineers, and researchers in speech aerodynamics modeling and articulatory speech synthesis. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3081496]

PACS number(s): 43.70.Aj [AL] Pages: 2272–2281

I. INTRODUCTION

The release burst and aspiration or "pop" associated with voiceless aspirated plosive consonants (e.g., /p^h/, /t^h/, and /k^h/) in many languages is a potentially important cue in the perception of these sounds[1] and a well-known challenge for audio engineers and microphone manufacturers.[2,3] Plosive release burst and aspiration contains both sound and what has been termed pseudo-sound.[4,5] While sound waves propagate through air at the speed of sound ($c=\sqrt{\gamma P/\rho}$ for an ideal gas), pseudo-sounds are slower pressure fluctuations within the flow that are detectable by an ear or microphone. The present paper seeks to characterize the properties of flow (as opposed to sound) associated with English aspirated "p."

While a good deal is known about the properties of air flow from an orifice in industrial applications, there exist a number of problems peculiar to modeling oral aspiration in speech that have not been previously addressed, including properties of the orifice, flow description, and simulation type.

A. Orifice

During the production of the labial plosive "p" release, the lips constitute a highly complex orifice, being elastic and continuously changing in geometry and rigidity. During English and Japanese bilabial stop releases, Westbury and Hashi[6] used Westbury's x-ray micro-beam data to demonstrate that the lips accelerate away from each other after the release, reaching a maximum velocity of about 200 mm/s at 25 ms, and then decelerate until they reach an average opening of ~20 mm after ~200 ms. While this mouth opening time is quick, it is not negligible compared to the time scales under study. Pelorson *et al.*[7] argued for the importance of modeling changes in the lip opening over time, but presumably did not simulate it due to computational complexity. The rate of lip opening is likely to have a large effect on initial flow rate as air flows faster through constrictions in a tube. However, disturbances due to interaction between the flow and lips are expected to have little effect on the general flow.[8,9]

In an engineering setting there have been few studies of the effects of variable orifice geometry on the fluid mechanics. One such study, by Dabiri and Gharib,[10] considered the starting jet formed by a circular orifice of time-varying diameter. They studied the effects of changing nozzle diameter on the flow and found that a temporally increasing nozzle diameter causes the leading vortex ring to have the strongest vorticity at a larger radius from the centerline than for a constant nozzle diameter, but they did not measure jet penetration distance, which is a primary quantity of interest here.

During the production of "pa's," lip aperture geometry is

a)Electronic mail: dderrick@interchange.ubc.ca

 0001-4966/2009/125(4)/2272/10/$25.00

close to an ellipse,[7] and so there is the need to consider whether modeling the general elliptical shape of the mouth opening is important in simulating airflow after a bilabial release. Noncircular jets have been previously studied as a means of providing passive flow control. Research results, both numerical and experimental, show significant differences between circular and elliptical jets;[11,12] thus simulating the general shape of the lip opening is likely to be important for accurate simulations.

B. Flow description

After a bilabial stop release into a vowel, the pressure in the mouth drops asymptotically to approximately 1/10th of its initial value in the first 60 ms of the flow (similar to Fig. 4).[7,13,14] The pressure in the mouth is sufficiently great that the flow out of the mouth during an utterance such as "pa" is turbulent. In a turbulent flow a large range of scales is present, as opposed to the smaller range present in a smooth, laminar flow. One can confirm that a "pa" is turbulent by considering the Reynolds number, which is a dimensionless parameter important for characterizing flows:

$$\mathrm{Re} = \frac{\rho V D}{\mu}.$$

In this expression, ρ is the fluid density, V is the mean fluid velocity at the orifice, D is the orifice diameter, and μ is the dynamic viscosity. For a Reynolds number greater than 1000, round jets become turbulent a short distance from the nozzle.[15] D is approximated as 6.1 mm by finding the hydraulic diameter for the mouth (see Ref. 14 for similar hydraulic diameters), $V=20$ m/s as a conservative estimate, and using typical values of air of $\rho=1.2$ kg/m^3 and $\mu=1.8\times10^{-5}$ N s/m^2, the Re is ~8100, so this flow is turbulent.

Turbulent starting jets (the initiation of a continuous flow from an orifice) and puffs (in which flow at the orifice is cut off soon after initiation) have been heavily studied for other applications such as fuel injection, and are typically studied with round nozzles. Sangras *et al.*[9] (note correction[16]) provided a nice summary of starting jet and puff research. The leading edge of the burst follows the equation (dropping the virtual origin)

$$X = cT^n,$$

where $X=x/D$ is the non-dimensional distance, c is an experimentally determined constant, $T=tV/D$ is the non-dimensional time, and $n=1/2$ for starting jets and $n=1/4$ for puffs. Figure 1 shows the difference between a starting jet and a puff using the range of c reported in Ref. 9. The puff and starting jet penetration distances diverge significantly for $T\gtrsim100$. Using the characteristic diameter and velocity of a "pa" estimated above, puffs and jets would penetrate noticeably different distances after about 30 ms. Since there is a need to understand "pa" behavior to 100 ms or more, it is clearly necessary to model the actual transient pressure driving the flow.

Assuming the room temperature to be 22 °C and the air jet to be 37 °C (body temperature), then from the ideal gas law one finds the ratio $\rho_0/\rho_{\mathrm{jet}}=1.05$. Diez *et al.*[17] found the effects of buoyancy to be small for the temporal and spatial range considered in this study, and, while Diez *et al.* considered buoyant forces acting along the streamwise direction, in speech the buoyant force will be roughly perpendicular to the jet, presumably resulting in an even smaller effect on streamwise penetration. Temperature will also cause the jet viscosity to be ~5% higher than the surrounding air, but this difference should also produce negligible effects on a flow at this Re.

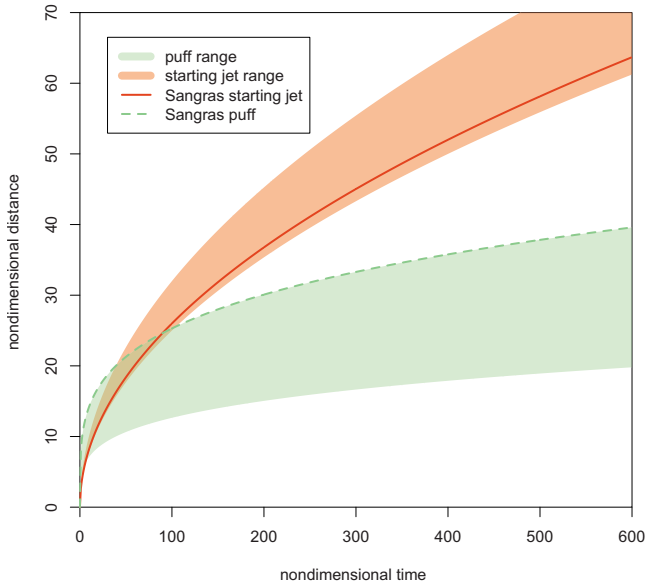


FIG. 1. (Color online) Starting puff and starting jet comparison.

C. Simulation type

One must consider whether the problem can be modeled in two dimensions, or if a more complex three dimensional (3D) model is required. If the domain is two dimensional (2D), then the mouth would have to be treated as a plane jet, as was done by Pelorson *et al.*[7] Although turbulence is a 3D flow, it is possible to consider a 2D Reynolds-averaged Navier–Stokes (RANS) turbulence model. However, Reichert and Biringen[18] and Stanley and Sarkan,[19] among others, reported significant inaccuracies in 2D simulations of plane jets. Finally, RANS models average the flow, but the turbulent fluctuations themselves are of interest to us; therefore a more sophisticated technique such as large eddy simulation (LES) is needed. Thus, both the geometry and the flow compel us to model plosive aspiration in three dimensions using LES.

D. Hypotheses

Based on the above discussion, it is proposed that to adequately simulate air flow from the mouth after the release of a bilabial stop into a vowel, one needs to take into account the known decrease in air pressure following the release. It is also hypothesized that the mouth can be adequately modeled as a 2D narrow ellipse. Computational limitations require a static geometry. The validity of this assumption must be compared with lip aperture over time from a high-speed video experiment. Due to the fact that the airflow throughout

most of the release is turbulent, it is necessary to resolve the turbulent properties, and because the lip aperture exists in 3D space, one needs a 3D LES to accurately model air flow after a bilabial release.

II. METHODS

These hypotheses were tested by comparing the results of two sets of experiments with simulations. The first experiment used a microphone located at varying distances from a participant repeating the syllable "pa" to record pressure fronts corresponding to microphone pops. The second experiment used high-speed video to record smoke particles. The microphone pops were compared to the simulation pressure front. The leading edge of the smoke particles recorded in the high-speed video was compared to the leading edge of the simulation particle front.

A. Microphone experiment

1. Data recording

For the microphone pop experiment, a single male participant was seated in a sound-proof room. Two microphones were placed in the room, one dummy microphone at 50 cm away from the mouth of the participant, and one SHURE SM58 set 5 cm away from the mouth of the participant. The cover of the microphone was removed to increase the effect of the pop on the recording, and the microphone was plugged into a Sound Devices USBPre microphone pre-amplifier plugged into a 1.42 Gbyte dual processor PowerMac G4 with 512 Mbytes of ram running Mac OSX 10.4.10 and recording with Audacity 3.3 at a sampling rate of 44 100 kHz. Both microphones were lined up and placed at exactly the mouth height of the participant.

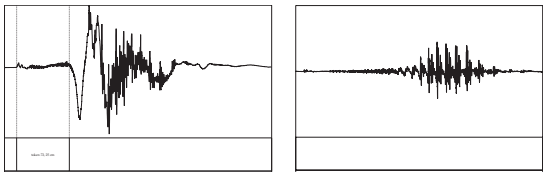
The participant wore a set of direct sound extreme isolation headphones plugged into the USBPre and set to monitor microphone input in real-time. The self-monitoring allowed the participant to adjust his speaking angle to make sure that microphone pops were being picked up by the Shure SM58 microphone, a particularly difficult task at distances past 20 cm.

The participant was handed a thin rigid tube to place in the corner of his mouth. The tube was attached to a SCICON Macquirer 516 airflow meter set to record the mouth pressure of the participant during the experiment. The airflow meter was attached to the same powerMac and using MACQUIRER 8.9.5.

The participant was asked to say the word "pa" 15 times while focusing on the dummy microphone set 50 cm away. The experiment was repeated with the microphone moved back at 5 cm increments from 5 to 40 cm away from the participant.

2. Data analysis

For each token the maximum air pressure just prior to the release burst of "pa" was recorded along with the difference in time from the onset of the sound of each "pa" and the beginning of a microphone pop. Airflow perturbations, or microphone pops, affect microphone output through the production of a very low frequency wave caused by the airflow, and high frequency aperiodic sound. To measure how long the airflow took to reach the microphone, the time between the onset of the release burst and the onset of the first significant low frequency perturbation that looks and sounds like microphone pop was used, as illustrated in token 75 in Fig. 2(a).



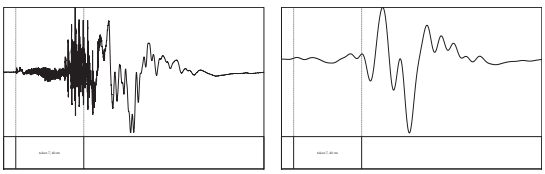
(a) 'pa' with pop (b) 'pa' without pop

FIG. 2. Sound waves from a "pa" with microphone pop and "pa" without microphone pop (183 ms clip).

However, these perturbations are difficult to isolate, particularly from a sound signal for distances from 20 to 40 cm due to overlap with the high amplitude vocalic portion of the sound wave. Fortunately, microphone pops are also associated with turbulence at higher frequencies. The high frequency aperiodic sound is hard to isolate in the waveform, but easy to detect by listening to the sound. Therefore each token was also examined by listening for the onset of pops using a set of high-quality Sennheiser HD650 headphones and a Total Bithead pre-amplifier. This turbulent sound helped isolate the onset of the microphone pop. For cases where neither listening nor examining the original wave worked, the original sound file was low-pass filtered using a band pass elliptic filter set from 30 to 100 Hz in MATLAB with 30 Hz skirts. These frequencies are produced in the sounds of speech, but microphone pops produce these frequencies at higher amplitude making the leading edge of the microphone pop easier to detect.

The time between the onset of the original sound wave and the onset of the first visibly larger peak was selected, but only when there was an obvious increase in the amplitude of these low frequency waves clustered together. This filtering method can reduce the accuracy of measurements because it excludes relevant frequencies that cannot be used because they overlap the fundamental frequency and first harmonic. However, in some cases the method was very helpful, as in token 7 shown in Fig. 3 where it is hard to see the onset of the pop in the unfiltered waveform, but easy to see in the low-pass filtered waveform.

If none of these three techniques produced a discernible result, the token was not used because the microphone did not record a loud enough pop to isolate.



(a) Unfiltered Sound Wave (b) Low Pass Filtered Sound

FIG. 3. Measurements from sound token 7, distance=40 cm, 437 ms clip.

The microphone pop timing corresponds to the leading pressure front recorded in the air puff simulation.

B. High-speed video experiments

Two sessions of high-speed video of the participant from the pop experiment saying the word "pa" while expelling white smoke were made. The smoke had a similar density as air, and was close to body temperature or approximately 37 °C at the time of expiration.

For the first round, digital videos of three productions of "pa" were captured using black foam board in the background and a standard tape measure pasted to the board for scale. The camera was placed approximately 460 cm away and focused on the tape measure such that the shot was 52.8 cm wide at the focal point. Bright sunlight was used to provide lighting. The participant then stood to the edge of the black bristol board such that their mouth opened just above the tape measure. The participant inhaled white smoke prior to the production of the "pa" so that the expelled air from the production of the "pa" would be visible during filming. Video was captured using a Bassler 504 kc high-speed color digital video camera with a Micro-Nikkor 70–180 mm telephoto zoom lens. The camera was plugged into an EPIX PIXCI CL3 SD frame grabber card with 1 Gbyte, of PC133 mHz memory in a P4 computer with 1 Gbyte of ram running Windows XP. Digital video was captured into the frame buffer using X-Cap Lite set to capture at 1024×768 resolution at 500 fps at maximum light gain and exported frame by frame into 1280×1024 32 bit tagged image file format (TIFF) files.

For the second round, digital video of 12 productions of "pa" was captured using black foam board background and meter sticks for scale. The camera was placed approximately 330 cm away and focused on the tape measure such that the shot was 53.0 cm wide at the focal point. A film light was placed facing the speaker to clearly illuminate the smoke particles. Video was captured using a Phantom v12 high-speed monochrome digital video camera with a Navitar 6.5 × lens. Digital video was transferred from the camera's built-in memory to 1280×800 resolution jpegs at 2000 fps.

1. Data analysis

For both rounds, the point of the opening of the mouth was captured using IMAGEJ's point capture utility, and the leading edge of the white smoke was recorded frame by frame for the first 150 ms of recorded time. The points were converted to distance in centimeters and analyzed statistically.

For both rounds, exact measurements of initial mouth pressure could not be made because the air flow apparatus would have interfered with the visual recording of air puff travel. However, the pressure can be inferred from Kenneth Stevens data on initial intra-oral air pressure during the production of aspirated stops at normal volume and the previous recordings of louder "pa's" during the microphone study which used the same subject (see Fig. 4 in Ref. 14).

For the second round, the rate of lip opening was also captured using IMAGEJ's point capture utility. The position of the top of the mucous membrane of the upper lip and the crease that intersects the mental protuberance and the skin below the lower lip were recorded frame by frame for 40 ms for each of the 12 recordings. These points provided stable landmarks for measuring the rate of lip opening. The points were converted to distance in millimeters and analyzed statistically.

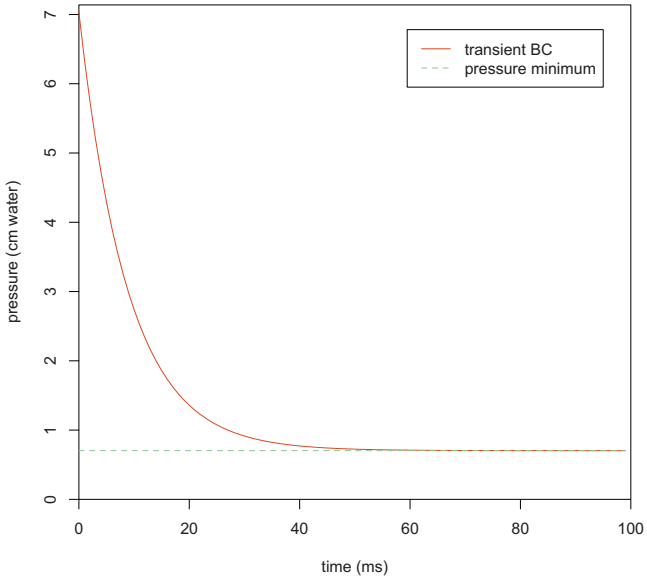


FIG. 4. (Color online) Transient boundary condition.

C. Numerical simulations

For the base numerical study, a domain of physical dimensions $350\times100\times100$ mm^3 which is meshed with 721 800, non-uniform, hexahedral control volumes was used. The mouth is shaped like a narrow ellipse in the $x=0$ plane, with $r_y=2$ mm and $r_z=15$ mm. A rough integration of upper and lower lip pellet velocities from the Westbury paper[6] shows that the lips have a y radius of 2 mm ~17 ms after they begin to separate.

Stevens[14] showed the intra-oral pressure quickly dropping after the release burst for "pa;" thus the mouth was modeled as a transient pressure inlet which quickly drops to 1/10th of its initial value, as shown in Fig. 4. In the simulation, the mouth lies in a plane that is modeled as a wall, while the rest of the boundaries are pressure outlets set to atmospheric pressure. The air is incompressible and initially still. Nitrogen particles were injected and tracked as a dye, thus defining the leading edge of the jet. An implicit bounded central differencing spatial discretization and an implicit second-order time discretization with a LES to model this turbulent flow were used. LES resolves the large eddies within the flow, but eddies smaller than the mesh scale are approximated by a turbulence model (in this case dynamic Smagorinsky[20]). The model was performed over 4000 time steps of size $\Delta t=0.025$ ms ($t_{\text{final}}=100$ ms). Using FLUENT as the solver, and running on three parallel processors, this process took ~6 days.

To explore the quality of the simulation methods and initial assumptions, numerous variations to this baseline

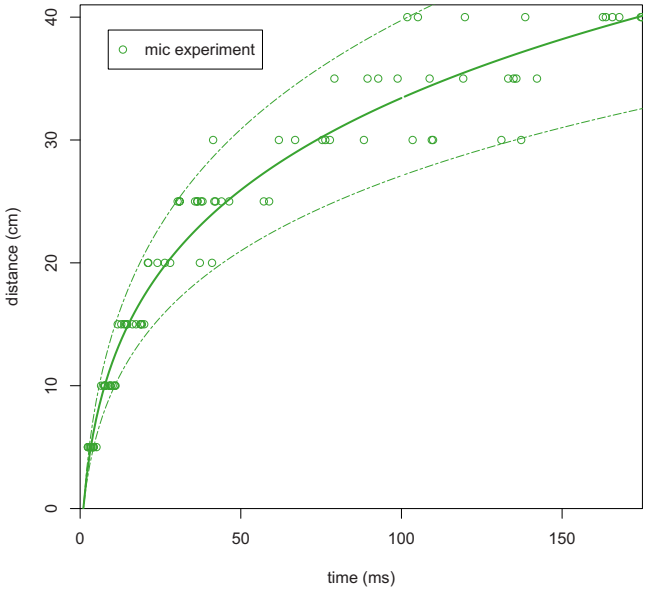


FIG. 5. (Color online) Experimental pressure fronts.

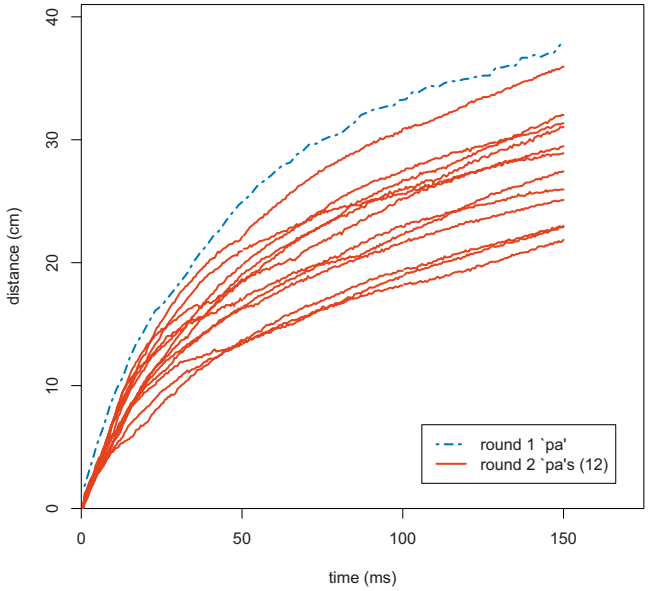


FIG. 6. (Color online) Leading smoke particle trails for pilot and second-round high-speed video experiment.

simulation were run; *Variation 1*: A grid refinement study was performed with the standard hexahedral mesh using a simulation with a medium mesh of 88 380 control volumes and a course mesh of 11 925 control volumes; *Variation 2*: A similar simulation replacing the mouth-shaped and time-varying inlet with a circular and constant velocity inlet, thus modeling a starting jet from a circular nozzle, was run in order to validate the numerical methods. See Ref. 21 for general discussion of verification and validation; *Variation 3*: A simulation with the starting inlet pressure three times higher than normal (24 cm/H_2O) yet falling to the same final value (0.703 cm/H_2O) was run to simulate a loud utterance: *Variation 4*: A simulation with a constant pressure inlet of 7.03 cm/H_2O (690 Pa), which is the same initial pressure of the baseline simulation was also run to test the importance of the transient pressure inlet; *Variation 5*: A simulation where the initial pressure was raised by 1 Pa was run. This slight change has little effect on the physics, but it does cause the numerics to change slightly, thus providing a second realization of the turbulent flow; *Variation 6*: A simulation was run where the inlet pressure condition was unchanged, but the initial domain was perturbed with small velocities, thus providing realistic disturbances in the air which are greater than machine zero. Some preliminary simulations in two dimensions, using LES and RANS were also conducted, but these soon proved to be inadequate.

III. RESULTS

The results of the microphone and high-speed video experiments, along with the numerical simulations, are described below.

A. Microphone experiment

Of 120 tokens recorded, 90 had discernible pops according to the standards described in Sec. II A. Individual measurements were highly variable, as seen in Fig. 5; the fit line is based on a loglinear quadratic fit with an assumed zero intercept. The fit line is highly significant, with an $F(2,88)=4386$, $p<0.001$ for each coefficient, and adjusted $R^2=98.9\%$. Linear, quadratic, cubic, and loglinear statistical models produced less significant results. As a result of trying to produce microphone pops in a microphone 50 cm away, the average intra-oral pressure was ~25 cm of water, or three times higher than normal, with high variability. This variability is largely a question of repeatability. It is almost impossible for a person to produce a repeatable mouth shape, initial air pressure, rate of decrease in air pressure, rate and degree of mouth opening, and orientation of the mouth to the microphone.

Many of these variables could not be measured and even initial mouth pressure could not be isolated from the other variables as no significant relationship was found between rate of air travel and intra-oral pressure prior to the release burst.

Nevertheless, the effect of many of these variables is known. Lower initial air pressure, faster rate of decrease in air pressure from the flow source, larger mouth opening, puff orientation away from the microphone, and perturbations in the air all decrease the rate of flow penetration. These effects combined can be quite significant.

B. High-speed video experiments

For the first round, three high-speed tokens were recorded, but only one was produced at a normal volume and voicing quality for an English "pa" syllable. This token was selected for comparison with the numerical simulations. For the second round of recordings, all 12 recordings were produced at a normal volume and voicing quality for an English "pa" syllable.

Results of measuring the leading edge of the smoke particles for each recording are shown in Fig. 6.

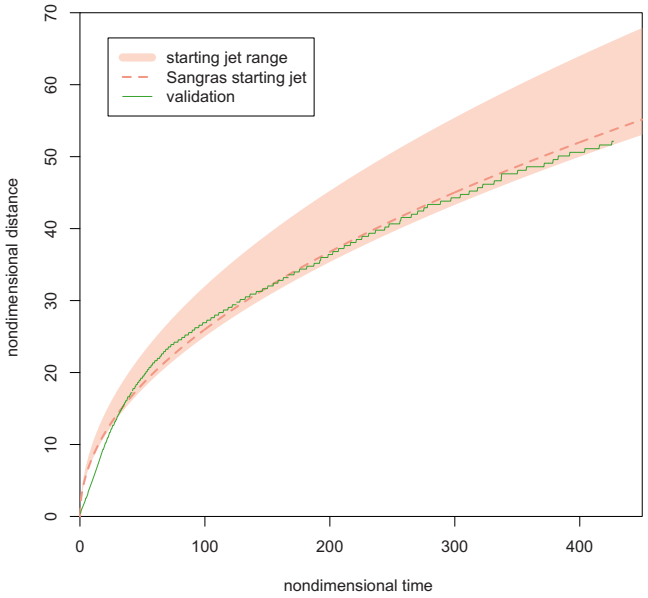


FIG. 7. (Color online) Numerical simulation validation: Starting jet range is defined by the constants reported in the summary of Sangras *et al.* (Ref. 9).

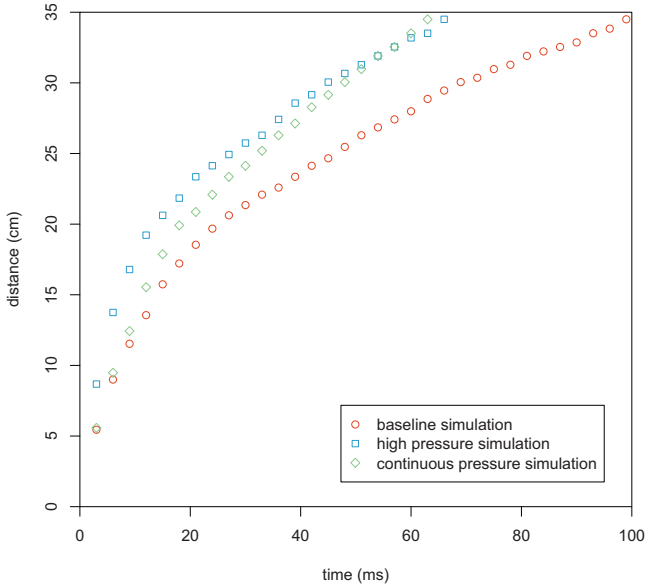


FIG. 9. (Color online) Comparison of leading particle front for baseline, high pressure onset, and continuous pressure simulation.

C. Numerical simulations

The validation study (variation 2) gives fine agreement with previous jet experiments described in the Introduction, as shown in Fig. 7. Figure 8 shows the grid refinement study (variation 1), along with the perturbed inlet simulation (variation 5) and the perturbed domain simulation (variation 6). The convergence is oscillatory, but outside of the asymptotic range. See Refs. 22 and 23 for discussion of oscillatory convergence and complications of LES verification, respectively. A comparison of the baseline numerical simulation, the simulation of the loud utterance (variation 3), and the constant inlet pressure simulation (variation 4) is presented in Fig. 9. As suggested in the Introduction, 2D simulations did not yield realistic results; generally they resulted in a jet penetration rate that was too fast. The loss of the 3D geometry caused the flow to be that of a plane jet rather than a jet from a nozzle. The loss of the 3D flow means that turbulence could not be truly modeled by LES, and the time-averaging of the RANS simulations removed flow details that are of interest. Use of 2D simulations was quickly dropped; therefore these results are not presented in detail here.

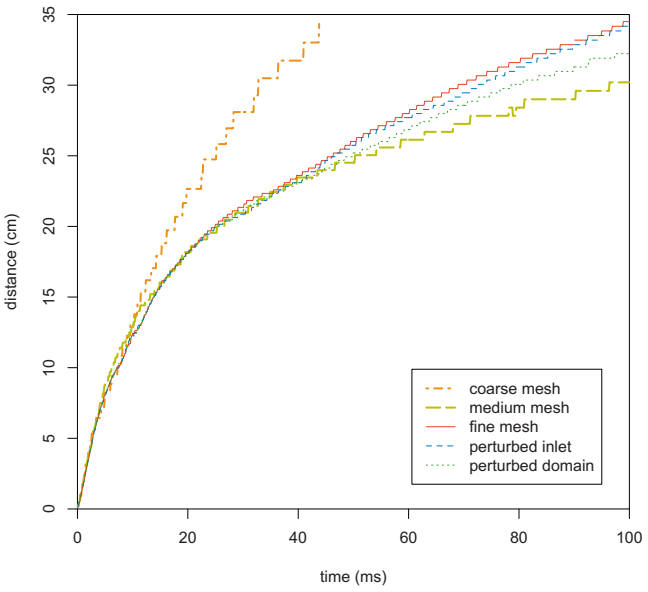


FIG. 8. (Color online) Simulation verification: Range represents confidence interval.

D. Comparison of simulation to microphone experiment

The simulation pressure front is defined as the distance at which the absolute value of the pressure reaches 1/10th of the maximum pressure for each time step. The simulation pressure front was compared to the results from the microphone experiment (Fig. 10). The green dots represent the mean measurements from the microphone experiment, the green line represents the loglinear quadratic fit, and the dashed green lines the 95% confidence interval. The simulation pressure front falls within the 95% confidence interval of the experiment.

E. Comparison of simulation to high-speed video experiment

A comparison graph between distance over time of the particle front from the high-speed video recordings and the numerical simulation appears in Fig. 11. The graph shows the loglinear quadratic fit lines for the pilot puff $[F(2,89)=1.125\times10^5,\ p<0.001,$ adjusted $R^2=99.9\%]$, second-round puff average $[F(3,3598)=7.479\times10^4,\ p<0.001,$ adjusted $R^2=97.7\%]$, and numerical simulation $[F(2,1208)=1.816\times10^6,\ p<0.001,$ adjusted $R^2=99.9\%]$. A comparison graph between the velocities over time of the particle front from the high-speed video recordings and the numerical simulation appears in Fig. 12. Note that the differences diminish dramatically after 40 ms, as shown in the inset within Fig.

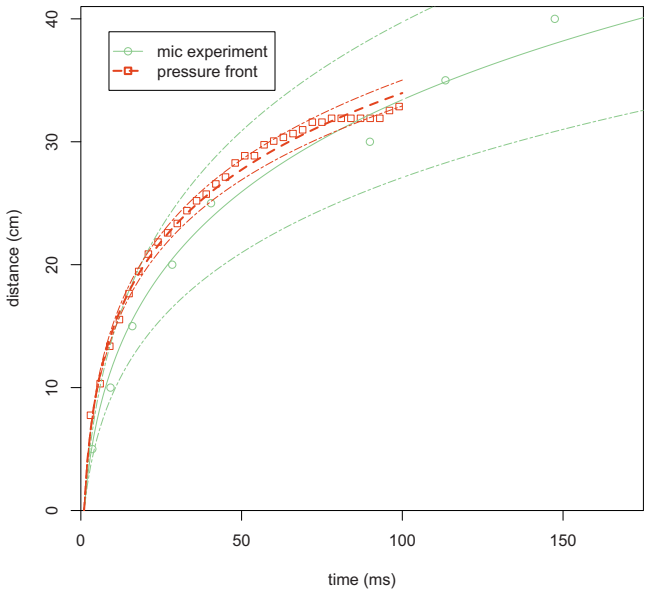


FIG. 10. (Color online) Average experimental and simulation pressure fronts.

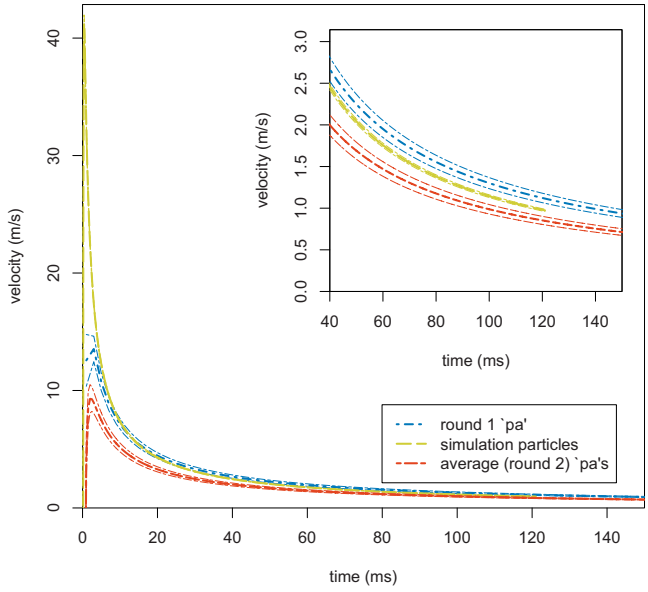


FIG. 12. (Color online) High-speed video particle front and simulation particle fronts (velocity).

12. There is a strong negative relationship between the rate of lip opening and leading particle edge distance traveled for the first 20 ms, diminishing after 30 ms and losing significance by 40 ms, as shown in Fig. 13. Both the significance and the *t*-value of the partial regression coefficient decrease over time as the leading edge of the puff moves away from the mouth opening. The results can be seen in Table I.

To illustrate the relationship between the smoke particle flow from high-speed film experiment and simulation results, a comparison of the video images from experiment round 1 and the baseline simulation is presented in Fig. 14. Round 1 video was selected to reduce the disparity between the images after time-alignment. Images were aligned such that the times at which the high-speed video's particle flow penetrates 5, 10, 15, 20, 25, 30, and 35 cm are matched with the same times in the simulation. Because frames are spaced 2 ms apart, the first frame with visible particle flow is assumed to occur ~1 ms after lip opening. This time-averaging, combined with the observation that the simulation flow rate matches closely, but not exactly with the high-speed video, creates distance alignment differences. As a result, the images do not align by particle front distance, and the differences can be seen in Table II. The velocity field instead of the particle field is shown because FLUENT does not export the particle data in a usable format and because the particle field can be inferred from the velocity field. In the high-speed video, most of the smoke is expelled in the first 30 ms, so the air expelled after that time is not as visible in the video frames. A graph of the simulated airflow velocity as a function of time in which each curve shows the velocity at a particular distance from the front of the orifice is shown in Fig. 15. The data are spatially averaged over a 1 cm radius in the *xz* plane and 2.1 ms in time. These lines reveal velocity oscillations around 100 Hz that were not smoothed out by the averaging. The oscillations are caused by large eddies in the flow that are resolved by the LESs, but which would not have been resolved with a RANS simulation.

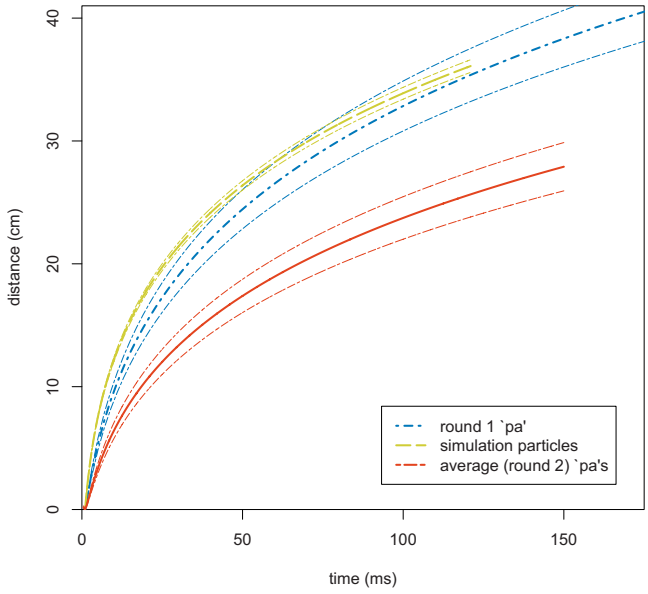


FIG. 11. (Color online) High-speed video particle front and simulation particle fronts (distance).

IV. DISCUSSION

These results show some significant discrepancies between the microphone experiment, the high-speed video experiment, and the simulations, but upon examination, these errors make sense in light of the assumptions and experimental methods used.

The microphone experiments were expected to show faster penetration than normal because the average intra-oral pressure was three times higher than normal, which was needed to attain good recordings. This impact, however, was not expected to be too large because velocity scales with the square root of pressure, as derived from Bernoulli's principle. Also, the air-pressure measurements showed that it

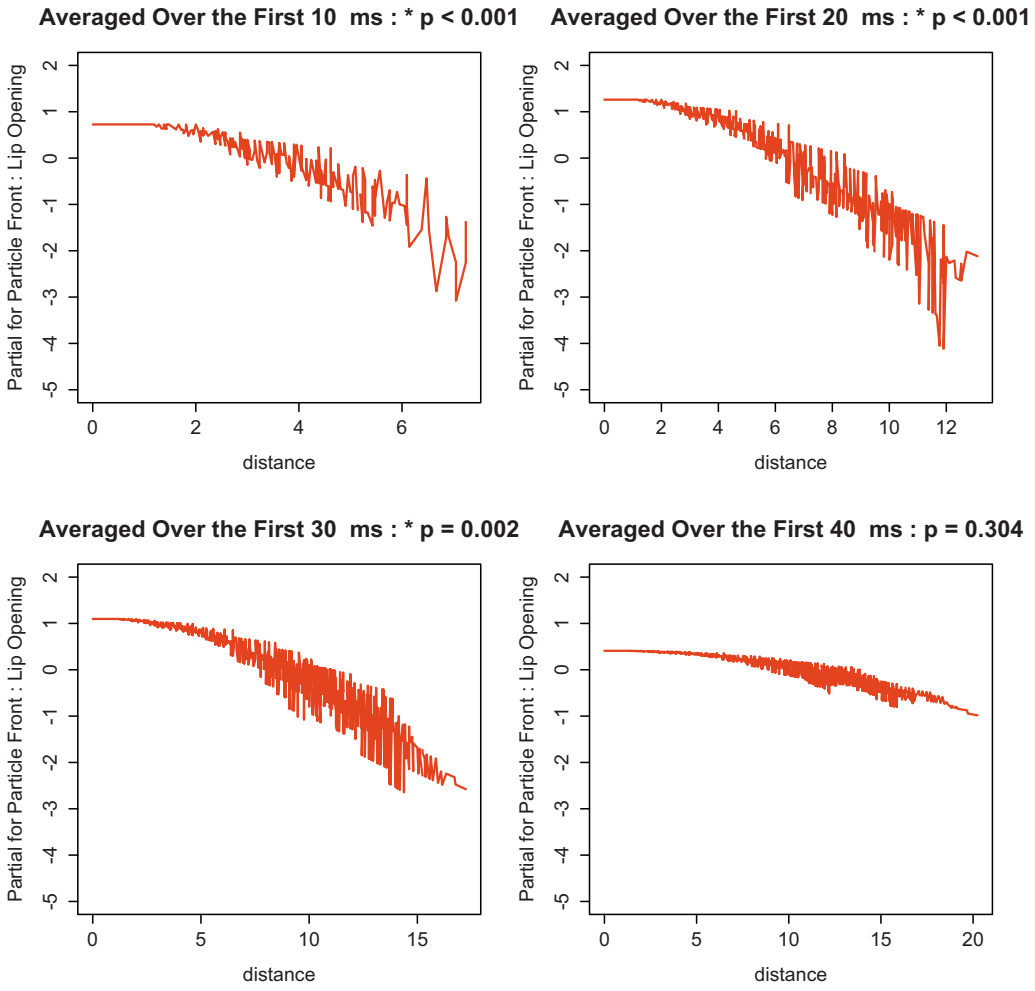


FIG. 13. (Color online) Negative partial regression between the width of lip opening and leading edge distance.

quickly fell to the normal level predicted in Stevens' book.[14] Therefore, the effect of the intra-oral air pressure would be less than one might expect, and the differences would be most significant at distances closest to the mouth.

The microphone experiment had a high variance due in part to the difficulty in capturing the microphone pops, especially at increasing microphone distances. Because of this high variance, the simulations fell within the range of results from the microphone experiment.

The high-speed video experiments, on the other hand, were captured at pressures reasonable for speech and were deemed trustworthy.

The measurements taken from the high-speed video were much more accurate than those taken from the microphone pop experiment as there were no visual artifacts interfering with visibility of the leading edge of the smoke comparable to the interference of the acoustic waves on the capture of microphone pops. Most of the variability in recorded results was seen in the rate of particle penetration during the first 40 ms, and could be largely attributed to the rate of lip opening during the first 20 ms. The faster the lips opened, the slower the initial penetration. Very tiny differences were significant.

TABLE I. Partial regressions of the interaction between the leading particle edge and lip opening averaged over 10, 20, 30, and 40 ms.

Coefficient	Time span (ms)	Estimate	Std. Err.	t	p
Puff travel	10	−1.69	0.34	−4.98	$*<0.001$
Distance	20	−0.81	0.18	−4.46	$*<0.001$
Lip opening	30	−0.36	0.11	−3.18	$*=0.002$
Width	40	−0.09	0.08	−1.03	0.304

The measured strong negative relationship between the rate of lip opening and the "pa" leading edge velocity was unexpected. Boundary layer effects would tend to produce a positive relationship between lip opening and "pa" velocity, so the negative relationship implies a more complex phenomenon, perhaps related to the geometry of the mouth behind the lips.

As discussed in the Introduction, there were a number of simplifying assumptions made for the simulation, particularly concerning the mouth. Not including the lips in the model means that the boundary layer effects were not modeled; these effects slow the jet. Also, Pelorson *et al.*[7] discussed the dominant role of viscous effects at the lips in the first milliseconds of a plosive, and Fujimura[24] also emphasized the rapidity of the change in the first 10 ms. Figure 12 shows that most of the simulation error occurs in the first 10–20 ms of the burst where the simulation velocity is much

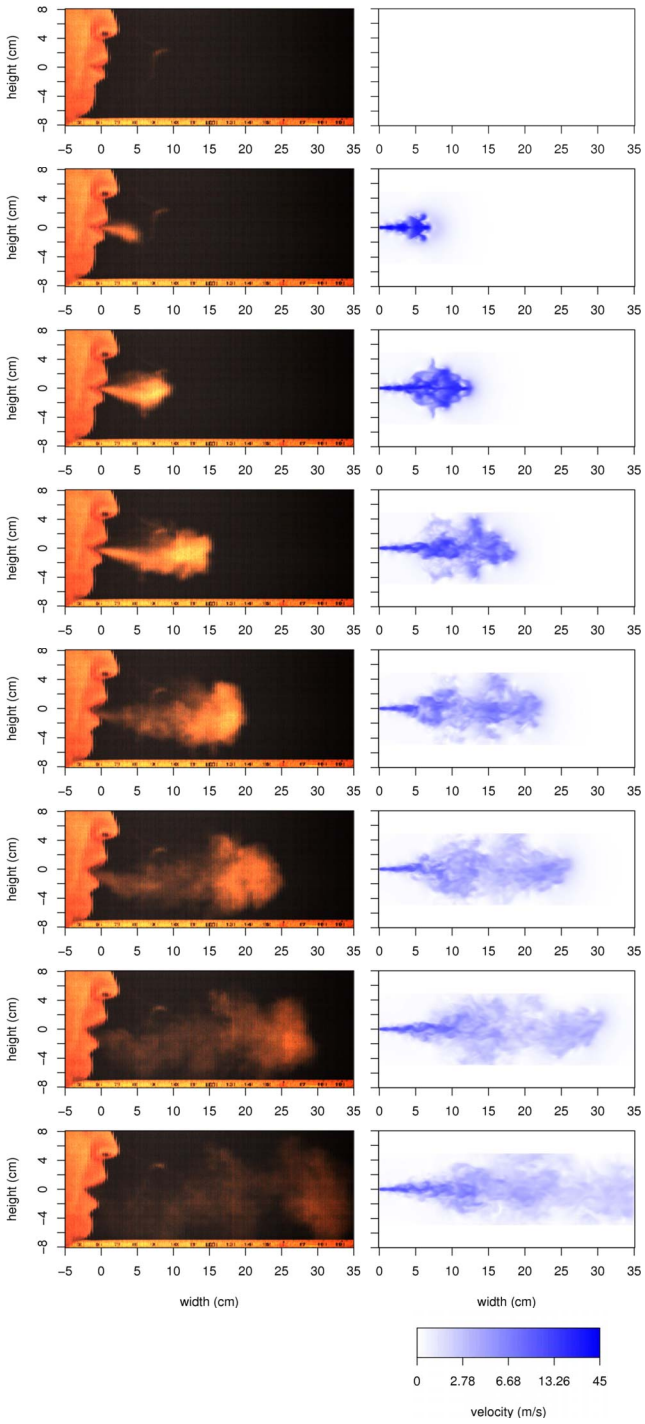


FIG. 14. (Color online) "Pa" on high-speed video (left) compared with numerical simulation velocity field (right). From top to bottom, the times (in milliseconds) of the image are 0, 5, 11, 21, 35, 51, 75, and 121.

higher than experiment, and the data in Fig. 13 and Table I confirm that the simulations suffer this error. This error largely accounts for the differences between the high-speed video experiment and the simulation.

A. Future work

The velocity data in this paper can be used to identify the maximum distance a perceiver can be positioned away from a speaker in order to detect puffs of air from labial plosives during their speech (though the minimum velocity at which skin receptors can detect air flow is as yet unknown), or as a basis for identifying the minimum distance a microphone needs to be from a speaker based on the microphone's sensitivity to air-flow velocity.

TABLE II. Time-alignment by distance for Fig. 14.

Time (ms)	Paticle distance (cm) by data source	
	High-speed video	Simulation
5	5.3	8.1
11	9.7	12.8
21	15.2	18.5
35	20.1	22.3
51	25.1	26.3
75	30.0	31.0
121	35.0	>34.8

While the simulations and the experiments match closely after 40 ms, the simulations predict faster airflow at the onset of the puff than that shown in the experiments. This difference was partially related to the fact that the mouth shape expands during the production of the "pa" syllable, but not in the simulation. Simulation of the change in oral aperture size would require changing the mesh throughout the simulation. This would be a challenging problem for further research. In addition, mesh and time step refinement may improve the quality of the simulations.

V. CONCLUSION

The results show that the hypotheses regarding the need for 3D LES with a mouth-shaped orifice and decreasing air pressure at the orifice are all reasonably valid for the accurate simulation of airflow after the release of an aspirated labial plosive. While the static elliptical orifice provided an adequate basis for simulation, the static and anatomically incorrect mouth shape contributed to the observed discrepan-

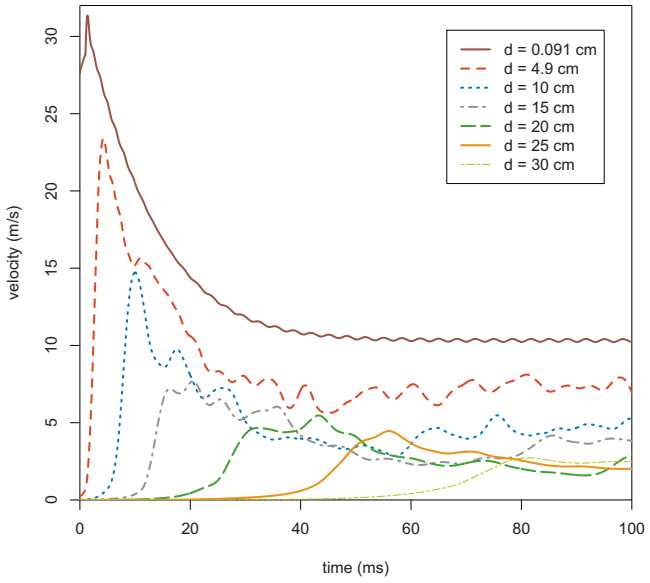


FIG. 15. (Color online) Airflow velocity over time based on the distance from orifice aperture.

cies in the results. Simulations involving a change in the orifice shape throughout the simulated time period, corresponding to known mouth shape changes in the production of labial plosives, may resolve this discrepancy.

By validating air-flow simulations to experimental data, it is possible to plot mean velocity in time as a function of downstream distance. This information can be used with experimental data to identify the distance away from the orifice or the time from the beginning of a speech release burst at which a person can perceive the airflow or a given microphone can pick up a pop.

These results provide the groundwork upon which future research in microphone manufacturing, sound engineering, speech perception research, and aerodynamic modeling of speech may be conducted.

ACKNOWLEDGMENTS

The authors thank Professor Sidney Fels and Professor Kees van den Doel for their advice and guidance, and Laurie McLeod and Walker Peterson for their help in conducting this research. They gratefully acknowledge support from NSERC grants to B.G. and S.G. and from NIH Grant No. DC-02717 to Haskins Laboratories.

[1]L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422 (1964).
[2]G. W. Elko, J. Meyer, S. Backer, and J. Peissing, "Electronic pop protection for microphones," in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2007), pp. 46–49.
[3]M. Schneider, "Transients in microphones: Pop and impulse," in Proceedings of the Audio Engineering Society UK Conference: Microphones and Loudspeakers (1998).
[4]M. J. Lighthill, "The Bakerian lecture, 1961. Sound generated aerodynamically," Proc. R. Soc. London, Ser. A **267**, 147–182 (1962).
[5]J. E. Williams, "Hydrodynamic noise," Annu. Rev. Fluid Mech. **1**, 197–222 (1969).
[6]J. R. Westbury and M. Hashi, "Lip-pellet positions during vowels and labial consonants," J. Phonetics **25**, 405–419 (1997).
[7]X. Pelorson, G. Hofmans, M. Ranucci, and R. Bosch, "On the fluid mechanics of bilabial plosives," Speech Commun. **22**, 155–172 (1997).
[8]S. C. Crow and F. H. Champagne, "Orderly structure in jet turbulence," J. Fluid Mech. **48**, 547–591 (1971).
[9]R. Sangras, O. C. Kwon, and G. M. Faeth, "Self-preserving properties of unsteady round nonbuoyant turbulent starting jets and puffs in still fluids," ASME J. Heat Transfer **124**, 460–469 (2002).
[10]J. O. Dabiri and M. Gharib, "Starting flow through nozzles with temporally variable exit diameter," J. Fluid Mech. **538**, 111–136 (2005).
[11]E. J. Gutmark and F. F. Grinstein, "Flow control with noncircular jets," Annu. Rev. Fluid Mech. **31**, 239–272 (1999).
[12]R. S. Miller, C. K. Madnia, and P. Givi, "Numerical-simulation of noncircular jets," Comput. Fluids **24**, 1–25 (1995).
[13]K. Stevens, "Airflow and turbulence noise for fricative and stop consonants: Static considerations," J. Acoust. Soc. Am. **50**, 1180–1192 (1971).
[14]K. Stevens, *Acoustic Phonetics* (MIT Press, Cambridge, MA, 2000).
[15]S. J. Kwon and I. W. Seo, "Reynolds number effects on the behavior of a non-buoyant round jet," Exp. Fluids **38**, 801–812 (2005).
[16]F. J. Diez, R. Sangras, O. C. Kwon, and G. M. Faeth, "Erratum: "Self-preserving properties of unsteady round nonbuoyant turbulent starting jets and puffs in still fluids" [ASME J. Heat Transfer, 124, pp. 460–469 (2002)]," ASME J. Heat Transfer **125**, 204–205 (2003).
[17]F. J. Diez, R. Sangras, G. M. Faeth, and O. C. Kwon, "Self-preserving properties of unsteady round buoyant turbulent plumes and thermals in still fluids," ASME J. Heat Transfer **125**, 821–830 (2003).
[18]R. S. Reichert and S. Biringen, "Numerical simulation of compressible plane jets," Mech. Res. Commun. **34**, 249–259 (2007).
[19]S. Stanley and S. Sarkar, "Simulations of spatially developing two-dimensional shear layers and jets," Theor. Comput. Fluid Dyn. **9**, 121–147 (1997).
[20]M. Germano, U. Piomelli, P. Moin, and W. H. Cabot, "A dynamic subgrid-scale eddy viscosity model," Phys. Fluids A **3**, 1760–1765 (1991).
[21]P. J. Roache, "Quantification of uncertainty in computational fluid dynamics," Annu. Rev. Fluid Mech. **29**, 123–160 (1997).
[22]I. Celik and O. Karatekin, "Numerical experiments on application of Richardson extrapolation with nonuniform grids," ASME J. Fluids Eng. **119**, 584–590 (1997).
[23]I. B. Celik, Z. N. Cehreli, and I. Yavuz, "Index of resolution quality for large eddy simulations," ASME J. Fluids Eng. **127**, 949–958 (2005).
[24]O. Fujimura, "Bilabial stop and nasal consonants: A motion picture study and its acoustical implications," J. Speech Hear. Res. **4**, 233–247 (1961).

Measurement of liquid and solid component parameters in canine vocal fold lamina propria

Robert Phillips, Yu Zhang, Megan Keuler, Chao Tao, and Jack J. Jiang
Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin-Medical School, Madison, Wisconsin 53792-7375

(Received 25 April 2008; revised 29 January 2009; accepted 30 January 2009)

This study aimed to measure solid and liquid component parameters for canine vocal fold lamina propria tissue, as is consistent with the solid and liquid fraction parameters in the context of the biphasic theory. A liquid-displacement apparatus was developed and utilized to estimate volumes of small samples of tissue. Accuracy was determined by calibrations with an object of known mass and density (copper). The experimental apparatus was then used to determine the volume of eight tissue samples, followed by an apparently complete dehydration of the samples, yielding the dry or solid tissue. The mass and volume fractions of the liquid component were sufficiently higher than those of the solid component. These results represent preliminary experimental evidence for the biphasic composition (solid-liquid) of canine lamina propria tissue as predicted in the biphasic theory. This study presents an effective experimental method to estimate some of the biphasic model parameters, and may provide a valuable application in exploring the viscoelastic behaviors of vocal fold lamina propria tissue. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3086276]

PACS number(s): 43.70.Bk, 43.35.Mr [BHS] Pages: 2282–2287

I. INTRODUCTION

Vocal fold (VF) tissue is known to be a highly hydrated interstitial tissue. The level of VF tissue hydration plays an important role in vocal function (Verdolini *et al.*, 2002). Studies have shown that VF dehydration can lead to an increase in phonation threshold pressure and a decrease in vocal efficiency (Verdolini *et al.*, 1994; Jiang *et al.*, 2000). Classic viscoelastic models of VFs (Chan and Titze, 2000; Chan and Tayama, 2002), such as the quasi-linear model and the statistical network model, have been applied to study viscoelastic behaviors of the tissue. However, these models do not account for the liquid dynamics in the VF tissue. Clinically, dehydration contributing to dysphonia is associated with the loss of liquid components of the tissue. Without accounting for the liquid component in VF tissue, studies have not quantified the effects of liquid loss on the biomechanical properties of VF tissue.

A biphasic theory has been introduced to describe the effect of liquid components of the VF tissue (Zhang *et al.*, 2008). VF tissue can be described as a porous solid component (collagen, proteoglycan, and other structural components) that is swollen by interstitial liquid. Biphasic (liquid-solid) descriptions of tissue, which consider the interaction between the solid and liquid components of the tissue, show potential for explaining rheologic properties of VF lamina propria (LP). It predicts that the liquid component of the tissue is critical for shielding the extracellular solid matrix from extreme stressors, indicating that dehydration may decrease the stress support offered by the interstitial liquid. In addition, the liquid component is important in determining the stiffness, viscosity, and mass of the tissue (Chan and Tayama, 2002). In the biphasic theory, the liquid volume fraction was defined to describe the ratio of liquid volume to the total tissue volume. The liquid loss due to dehydration can be quantitatively measured by a decrease in the liquid volume fraction. Therefore, the direct measurement of the liquid and solid components is necessary for the application of the biphasic theory, a better understanding of the multiphasic composition of VF tissue and the effect of these parameters on biomechanical tissue properties. Despite these parameters' utility, liquid and solid volume fractions of laryngeal systems have not been measured in previous studies, although analogous parameters have been quantified in other biomedical systems, such as articular cartilage (Mow *et al.*, 1980, 1984).

The purpose of this study was to measure the individual contributions of the liquid and solid components to the total VF LP tissue's mass and volume. Due to the small tissue samples, we will apply a liquid-displacement apparatus to measure tissue volume. We will examine the accuracy of the experimental apparatus through measurements of copper samples, which has a known density and constant volume. Tissue volume measurement, in combination with the dry mass, allows for the calculation of the mass and volume fractions of the liquid and solid in the tissue. This information may lead to a better understanding of the biphasic description of VF LP tissue.

II. MATERIALS AND METHODS

A. Tissue sample preparation

Four larynges were harvested from healthy laboratory dogs, which yielded the eight tissue samples used in the experiment. Excision of the larynx was performed according to the procedure described by Jiang and Titze (1993). The excised larynges were placed in a 0.9% saline solution with a calculated density ρ_{sa} of 1.004 g/ml (McCutcheon *et al.*, 1993). The larynges were flash frozen in liquid nitrogen, stored in a freezer at −20 °C, and slowly unthawed in saline

 0001-4966/2009/125(4)/2282/6/$25.00

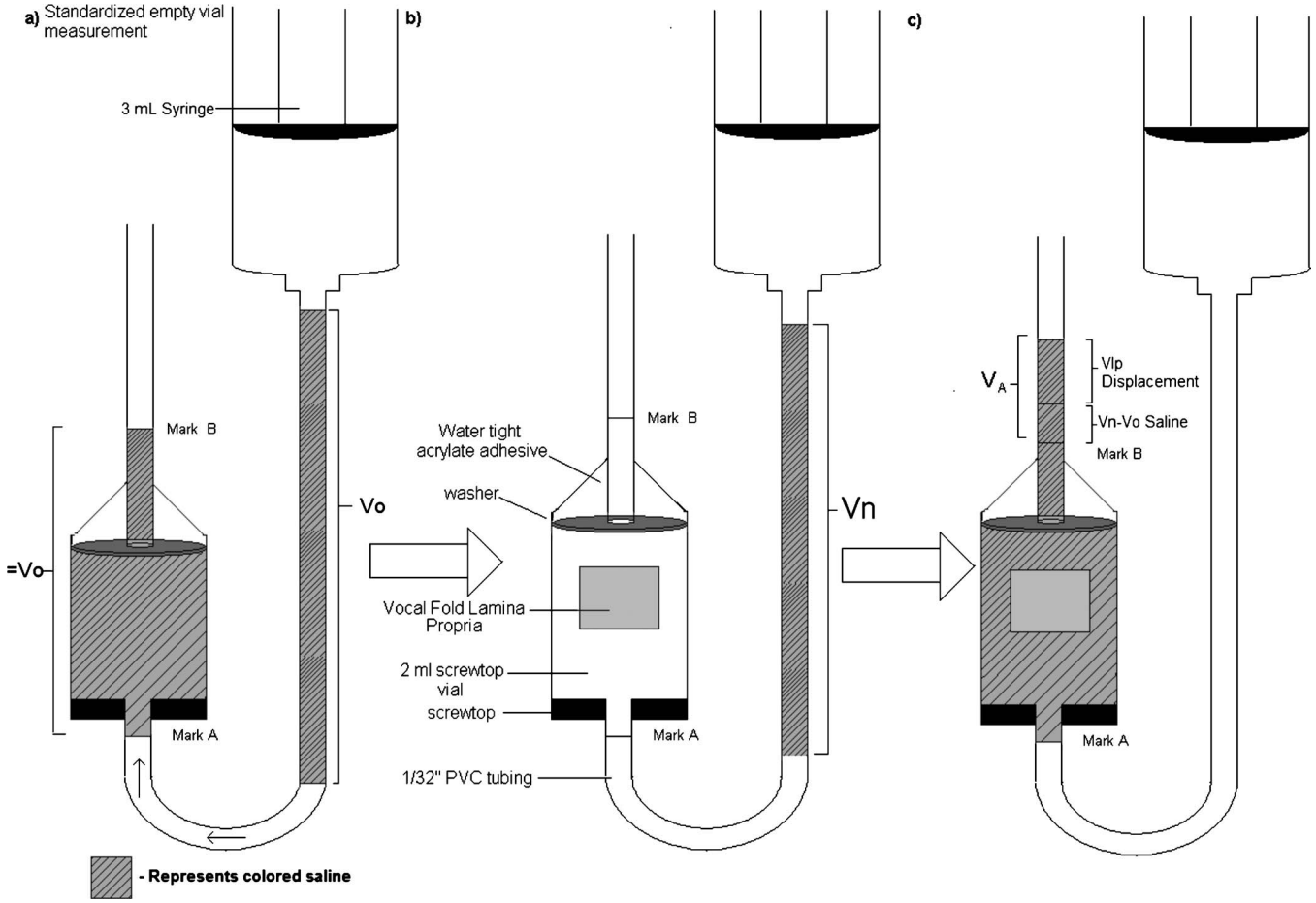


FIG. 1. Schematic of the liquid-displacement measurement process.

at about 4 °C a day before the experiment according to the Chan and Titze (2003) procedure which was found to minimize *postmortem* changes to VF tissue. A VF LP tissue was dissected out en block using a scalpel, with the anterior border 1 mm posterior to the anterior commissure, the posterior border 1 mm anterior to the arytenoid cartilage, the inferior border 5 mm inferior to the superior glottal edge, and the lateral border 5 mm lateral to the glottal edge. The thyroarytenoid muscle was then dissected away from the LP under a dissecting microscope using a scalpel. The epithelium was included in each sample in order to preserve the entire superficial lamina propria. The sample was then placed in a beaker containing 0.9% saline and allowed to equilibrate for 5–10 min before further handling.

B. Liquid-displacement method

Due to the small size of VF tissue, directly measuring tissue volume would bring sufficient errors. We applied the liquid-displacement method to measure the LP tissue volume V_{LP}. Figure 1 shows a schematic of this method, where the apparatus was assembled using approximately 2 m of 1/32 in. diameter clear PVC tubing, a 2 ml screw-top vial, a washer, silicone, and an acrylate adhesive. The screw-top vial was inverted and connected vertically downward (through the cap) to approximately 1.5 m of tubing in a U-bend shape. About 0.5 m of tubing was then run through the top of the vial using silicone adhesive for sealant. A washer was placed inside the vial, wedged against the chamber sides and below the silicone adhesive, preventing any volume changes in the chamber due to silicone compression.

Colored 0.9% saline solution was pushed via a 3 ml syringe (Terumo Supply, Elkton, MD) into the upper portion of the 1.5 m section of tubing and the height of this column was measured with an accuracy of 0.01 cm. 50 ml portions of saline had been colored using 1 drop of food dye to aid in measuring saline column height and identifying residual liquid remaining in the tubing. It was assumed that the small amount of dye used did not significantly affect the density or osmotic characteristics of the saline solution. Therefore, the density of the solution was calculated at 1.004 g/ml (McCutcheon *et al.*, 1993). Utilizing the known diameter of the tubing and the height of the water column, we were able to calculate the volume of this solution, V_0. The syringe was attached to the upper 1.5 m section throughout the procedure and facilitated the movement of the saline column along the tubing by "pushing" air into the system. This known volume was then pushed through the tubing so that it was approximately 1 cm below the vial, completely filling the vial, and a few centimeters above the vial. The tubing was continuously checked throughout the process to assure that the least amount of residual liquid possible was left behind in the tubing. According to the ratio of residue to total volume pushed through the tubing, it was estimated that far less than 1% of the original volume remained adhered to the inner

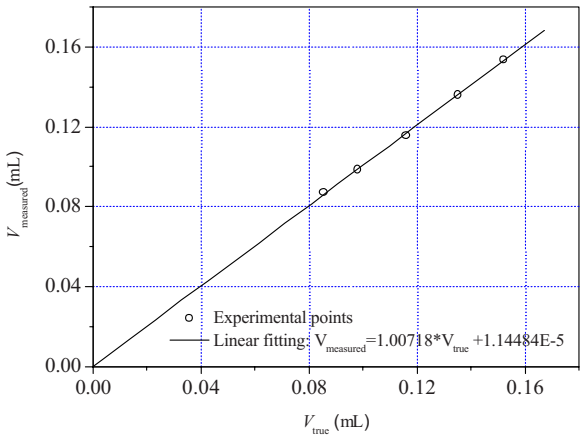


FIG. 2. (Color online) The relationship between the actual volume (V_{true}) and the experimentally measured volume ($V_{measured}$) of the five small copper samples.

tubing surface. The levels of solution below and above the vial were marked as marks A and B, respectively, and were considered to contain the known volume, V_0, between them. The fluid levels between marks A and B were measured to a ± 0.5 mm accuracy, which reflects an error value even less than that caused by the above saline-residue error. This volume was considered the "empty vial" volume since this part of liquid was taken out and was used to determine the volume of LP tissue in successive experiments, as shown in Fig. 1(a).

The tissue sample was then weighed on an electronic balance (SPE123, Ohaus Corporation, Pine Brook, NJ) accurate to 0.001 g, and immediately placed into the vial, soon to contain the new volume of saline, V_N. A new volume of liquid (V_N) was pumped into the upper 1.5 m of tubing, and the volume of this solution was calculated by measuring the height of the new column, and utilizing the known diameter of the tubing, as shown in Fig. 1(b). This volume was then pushed through the vial containing the excised LP tissue until the lower level of the fluid was exactly level with mark A. The liquid level above mark B was measured and used to compute V_A, or the volume of saline above mark B [Fig. 1(c)]. V_A includes the contribution from the tissue volume (V_{LP}) and the volume difference of $V_N - V_0$. From the measured V_0, V_A, and V_N, we can thus obtain the tissue sample volume V_{LP} by using the relationship

$$V_{LP} = V_0 + V_A - V_N. \quad (1)$$

In order to examine the accuracy of the volume measurement by liquid-displacement method, we used pieces of 99.99% pure copper with the density of 8.92 g/ml as test samples. Five small pieces of copper, all of which fit into the vial without restricting the flow of saline, were weighed with an accuracy of 0.001 g. From their mass and density, we calculated the volumes V_{true} of these copper pieces. The experimental volumes $V_{measured}$ of these copper pieces were measured using the liquid-displacement method. The volume measurement of each copper sample was repeated five times. Figure 2 shows the relationship between V_{true} and the averaged $V_{measured}$ of these copper pieces. The replicated volume measurements for all copper samples have the maximal error of 1.75% from their true values. The measured volumes V_{true} are a very close fit to a linear trend ($R=0.99964$), indicating a highly accurate volume measurement of small copper pieces.

C. Measurement of the volume and mass fractions of liquid and solid in VF lamina propria tissue

After measurement of the tissue volume V_{LP}, the mass of the tissue sample was measured on the electronic balance. The LP includes both liquid and solid components; therefore, its total hydrated mass was recorded as m_{LP}. The tissue density can be determined as $\rho_{LP}=m_{LP}/V_{LP}$. In order to completely dehydrate the tissue, the tissue samples were left in a vacuum oven (Isotemp Model 280A, Thermo Fisher Scientific, Waltham, MA) heated to 40 °C, and with the pressure lowered to approximately −15 mm Hg. The tissue was kept in the oven until it maintained a constant mass, which was denoted m_s or the dried solid component of the lamina propria. The difference between the total mass of the hydrated tissue and the dry solid component mass represents the loss of liquid from the tissue, allowing the mass of the liquid component to be computed as $m_l=m_{LP}-m_s$. We define the solid mass fraction ϕ_m^s as the fraction of the total mass represented by the solid mass and the liquid mass fraction ϕ_m^l as the fraction of the liquid mass of the total tissue mass. From the measured m_s and m_{LP}, we then can determine ϕ_m^s and ϕ_m^l as

$$\phi_m^s = \frac{m_s}{m_{LP}} \times 100\% \quad \text{and} \quad \phi_m^l = \frac{m_{LP} - m_s}{m_{LP}} \times 100\%. \quad (2)$$

Furthermore, the porous VF LP tissue was immersed in the physiological saline, which mimics the density and osmolarity of liquid in the tissue. The physiological saline as a bathing solution is considered to be isotonic to the tissue, thus preventing a net osmotic diffusion of water into or out of the tissue. Since the tissue can be viewed as in equilibrium with the saline, we can assume that the liquid density ρ_f sufficiently approached the density 1.004 g/ml of the saline solution. We then can obtain the liquid volume as $V_l=(m_{LP}-m_s)/\rho_f$. The difference between the total volume and the fluid volume gives the solid volume as $V_s=V_{LP}-V_l$. Similarly, we define the solid volume fraction ϕ_V^s as the fraction of solid volume versus total volume and the liquid volume fraction ϕ_V^l as the percent of the total volume occupied by the liquid phase. Thus, from the measured m_s, m_{LP}, and the total volume V_{LG} of the lamina propria, we can determine ϕ_V^s and ϕ_V^l as

$$\phi_V^s = \frac{V_{LP} - (m_{LP} - m_s)/\rho_f}{V_{LP}} \times 100\% \quad \text{and}$$

$$\phi_V^l = \frac{m_{LP} - m_s}{\rho_f V_{LP}} \times 100\%. \quad (3)$$

Clearly, these parameters are defined such that $\phi_m^l + \phi_m^s = 1$ and $\phi_V^l + \phi_V^s = 1$. During the experiment, the volume measurement was repeated on each tissue sample three times to ensure consistency in computing these parameters. The average variation of these measurements was 2.85% of the total tis-

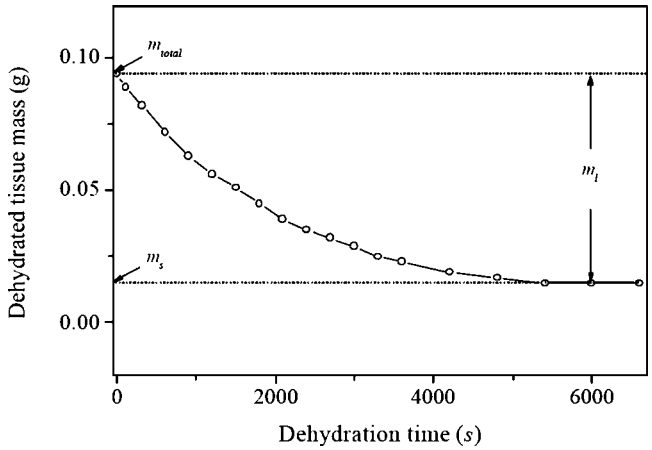


FIG. 3. The estimated solid and liquid masses of a VF LP tissue during dehydration.

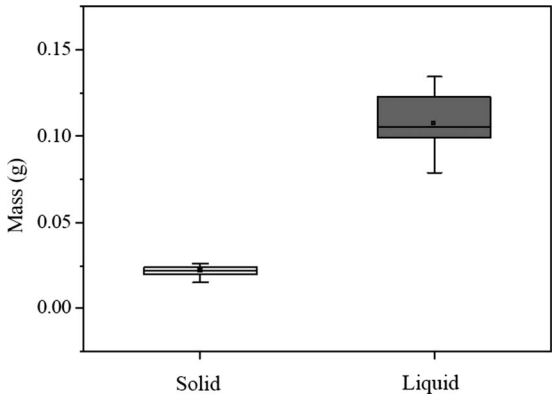


FIG. 5. The distributions of the solid and liquid masses from eight tissue data pieces excised from four larynges, where the line inside the box marks the median, and the upper and lower lines show maximum and minimum values. Error bars represent 10th and 90th percentiles.

sue volume. The masses (m_l, m_l) and volumes (V_s, V_l) of the eight VF LP samples from four excised larynges were measured in order to compare the results of liquid and solid components for each parameter.

III. RESULTS

Figure 3 shows the typical results of the solid and liquid masses of a VF LP tissue. Before dehydration, using the electronic balance and liquid-displacement method, the total mass m_{LP} and volume V_{LP} of the LP tissue were measured as 0.094 g and 0.090 m, respectively. The tissue density ρ_{LP} can be estimated as 1.044 g/m. After a thorough dehydration of the tissue, its mass approached a constant, as shown in Fig. 4, and thus the solid and liquid masses can be obtained as m_s=0.015 g and m_l=0.079, respectively. The liquid and solid volumes were estimated as V_l=0.079 ml and V_s =0.011 ml, respectively. Figures 4(a) and 4(b) illustrate the tissue sample before and after the complete dehydration. It can be seen that the dehydration sufficiently decreased the tissue volume. Based on Eqs. (2) and (3), we can determine the solid and liquid fraction parameters of this tissue as ϕ_m^s =16.0%, ϕ_m^l=84.0%, ϕ_V^s=12.2%, and φ_V^l=87.8%.

Using the same procedure, we obtained the solid and liquid fraction parameters of all eight tissue samples harvested from the four larynges used for the experiment. Figure 5 shows the distribution of the solid mass m_s and the liquid mass m_l. Mean values of m_s and m_l were estimated as 0.023 g and 0.111 g. Figure 6 shows the distributions of the solid volume V_s and the liquid volume V_l. The liquid volume with the mean value 0.111 ml was sufficiently higher than the solid volume with the mean value of 0.017 ml. Based on Eqs. (2) and (3), the fraction parameters ϕ_m^s, ϕ_m^l, ϕ_V^s, and ϕ_V^l can be obtained, as summarized in Table I. The mean value of the liquid mass fraction ϕ_m^l was 83.0%, which is sufficiently higher than that of the solid mass fraction ϕ_m^s (17.0%). The liquid volume fraction ϕ_V^l of 86.7% was higher than the solid volume fraction ϕ_V^s of 13.3%.

FIG. 4. (Color online) A tissue sample before (a) and after (b) dehydration.

IV. DISCUSSION

Past studies on articular cartilage have used the buoyancy (sink-float) procedure to measure the density of small tissue samples, which can be used to calculate volume (Lipshitz *et al.*, 1976). However, this procedure requires an equilibrated solution of specific concentration, *p*H, and temperature, making it difficult to control for the permeable nature of the VF tissue in varied osmotic environments. Since the tissue would need to be submerged in a solution for this sink-float procedure, there may be a net diffusion of water, making an accurate volume measurement difficult. Other

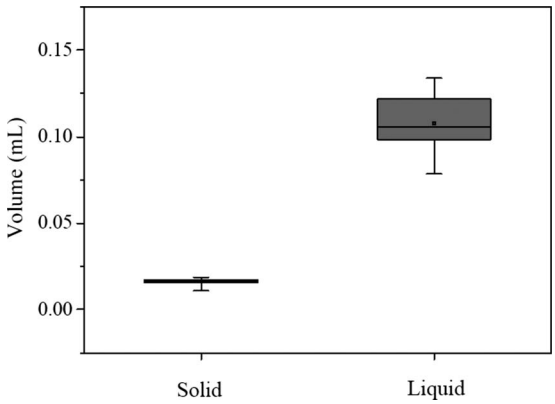


FIG. 6. The distributions of solid and liquid volumes, from eight tissue data pieces excised from four larynges, where the line inside the box marks the median, and the upper and lower lines show maximum and minimum values. Error bars represent 10th and 90th percentiles.

TABLE I. Comparisons of solid and liquid groups on mass and volume fractions from eight tissue data pieces excised from four larynges, where the parameter values are given for each group (mean ± standard deviation).

	Solid	Liquid
Mass fraction ϕ_m (%)	17.0 ± 0.92	83.0 ± 0.92
Volume fraction ϕ_V (%)	13.3 ± 1.46	86.7 ± 1.46

procedures used to measure large tissue volumes include ultrasound technology (Hornblower *et al.*, 2007) and magnetic resonance imaging (Steen *et al.*, 2007), but these require relatively expensive equipment, and may be inconvenient for the measurement of VF samples with a small volume. To overcome the limitations of these methods, we have applied a liquid-displacement method to directly measure the small tissue volume of LP tissue. The current apparatus (Fig. 1) provides an inexpensive and simple way to measure small tissue samples. Differing from the buoyancy procedure, the liquid-displacement method allows the properties of the tissue to be conserved in a physiologic solution. In addition, as shown in Fig. 2, the linear relation and small deviation between the actual V_{true} and measured V_{measured} copper volumes showed the accuracy of the liquid-displacement apparatus and methodology. The precision of the apparatus in measuring copper samples allowed us to take the volume measurements in samples as small as 0.085 ml.

We measured the mean density of eight LP tissue samples as 1.049 g/ml, which is close to the density of the vocalis muscle reported in investigations by Perlman *et al.* (1984). Such a close relationship between LP density and vocal muscle density has also been used in recent studies of ultrasonic measurement (Huang *et al.*, 2007) and computer modeling of VF tissue (Titze and Hunter, 2004; Jiang *et al.*, 2001; Alipour *et al.*, 2000). The relative density difference $|\rho_{\text{sa}}-\rho_{\text{LP}}|/\rho_{\text{LP}}$ between the saline solution and the LP tissue was less than 4%. This implies that errors in determining the liquid density ρ_f and other fraction parameters may be less than 4%. However, error values this small are much less than the differences between the liquid and solid fraction parameters, since the liquid fraction parameters are at least five times larger than the solid fraction parameters, as shown in Fig. 6.

Limitations in our study resided in the number and type of larynges used. Our preliminary estimations are based on eight tissue samples from four distinct larynges which is not a large enough data set. Yet these estimations are still of value since they indicate a substantial difference in the currently assumed liquid and volume fractions. However, a greater number of laryngeal measurements will be included in our future studies. Another limitation in our study is that our *in vitro* estimations may not be the best representation of the dynamic relationship between the fluid and solid phases of the *in vivo* laryngeal system.

The VF LP tissue can be thought of as a porous solid structure swollen by fluid. The solid structure is comprised of collagen, elastin, and proteoglycans. Collagen is a fibrous structural protein and elastin is the other main structural protein. Highly branched proteins, called proteoglycans, fill a large amount of the superficial layer juxtaposed with the other fibrous proteins. The fibrous units form a cohesive porous composite solid phase in VF lamina propria. VF solid structure allows for a large amount of liquid storage and permeable flow. For example, hyaluronic acid is one component that can attract and hold water within the LP forming large, space-filling aggregates. Tissue solid components interact with the liquid regulating its distribution and flow. Homogeneous, solid descriptions of the VF tissue have primarily been applied in previous lumped mass and continuum models (Titze and Hunter, 2004; Jiang *et al.*, 2001; Alipour *et al.*, 2000). These single phase solid models are inadequate in describing the intrinsic solid-liquid interaction in the VF tissue and phenomena such as tissue stress relaxation. The consideration of the liquid contribution in VF models may be important for better understanding the viscoelastic behavior of VF lamina propria.

We recently proposed a biphasic model of VF LP tissue to describe the solid-liquid interaction (Zhang *et al.*, 2008). The biphasic theory predicts the stress relaxation of the LP and the crucial role of the liquid in stress support. However, there is no published research on the solid-liquid ratio of VF tissue, so the solid and liquid volume fractions (ϕ_V^l=80% and ϕ_V^s=20%) from the articular cartilage system (Mow *et al.*, 1980, 1984) were used in that study. In this study, by developing an apparatus which allows the volume measures of solid and liquid components, we provided preliminary experimental measurements of the biphasic (solid-liquid) parameters of LP tissue. In Table I, the mean liquid volume fraction ϕ_V^l and solid volume fraction ϕ_V^s of eight tissue samples harvested from four larynges were estimated as 86.7% and 13.3%, respectively. The lower ϕ_V^l in articular cartilage may be associated with differences in solid component composition. The amount or distribution of solid phase components holding water in the tissue could relate to differences in articular and VF ϕ_V^l. Biomechanical properties of VF tissue are fundamental factors affecting VF vibration and voice production (Chan *et al.*, 2001). Our study investigated the biomechanical property of VF LP tissue by measuring its biphasic parameters, which may provide useful information for understanding the VF vibratory systems.

V. CONCLUSION

In this study, we have measured the liquid and solid component parameters in canine VF lamina propria. A volume displacement apparatus was used to determine the volume of the VF LP tissue. The volume measurement of small copper samples examined the accuracy of the experimental apparatus. Using eight LP tissue samples, we found that the mass and volume of the liquid component in VF LP tissue were sufficiently higher than those of the solid component. Average mass and volume fractions of the liquid component were estimated as 83.0% and 86.7%, respectively, which are sufficiently higher than 17% and 13.3% of the solid mass and volume fractions. Liquid represents the major composition of the VF LP tissue. The results showed preliminary experimental evidence for the biphasic property (solid-liquid) of LP tissue. The estimated biphasic parameters may serve as cru-

cial parameters for future theoretical as well as experimental investigations exploring the viscoelastic behaviors controlling the vibratory motion of VF LP tissue.

ACKNOWLEDGMENTS

This study was supported by NIH Grant No. 1-RO1DC05522 from the National Institute on Deafness and other Communication Disorders.

Alipour, F., Berry, D. A., and Titze, I. R. (**2000**). "A finite-element model of vocal-fold vibration," J. Acoust. Soc. Am. **108**, 3003–3012.

Chan, R. W., and Titze, I. R. (**2003**). "Effect of postmortem changes and freezing on the viscoelastic properties of vocal fold tissues," Ann. Biomed. Eng. **31**, 482–491.

Chan, R. W., Gray, S. D., and Titze, I. R. (**2001**). "The importance of hyaluronic acid in vocal fold biomechanics," Otolaryngol.-Head Neck Surg. **124**, 607–614.

Chan, R. W., and Tayama, N. (**2002**). "Biomechanical effects of hydration in vocal fold tissues," Otolaryngol.-Head Neck Surg. **126**, 528–537.

Chan, R. W., and Titze, I. R. (**2000**). "Viscoelastic shear properties of human vocal fold mucosa: Theoretical characterization based on constitutive modeling," J. Acoust. Soc. Am. **107**, 565–580.

Hornblower, V. D. M., Yu, E., Fenster, A., Battista, J. J., and Malthaner, R. A. (**2007**). "3D thoracoscopic ultrasound volume measurement validation in an *ex vivo* and *in vivo* porcine model of lung tumours," Phys. Med. Biol. **52**, 91–106.

Huang, C., Sun, L., Dailey, S. H., Wang, S., and Shung, K. K. (**2007**). "High frequency ultrasonic characterization of human vocal fold tissue," J. Acoust. Soc. Am. **122**, 1827–1832.

Jiang, J., Verdolini, K., Aquino, B., Ng, J., and Hanson, D. (**2000**). "Effects of dehydration on phonation in excised canine larynges," Ann. Otol. Rhinol. Laryngol. **109**, 568–575.

Jiang, J. J., and Titze, I. R. (**1993**). "A methodological study of hemilaryngeal phonation," Laryngoscope **103**, 872–882.

Jiang, J. J., Zhang, Y., and Stern, J. (**2001**). "Modeling of chaotic vibrations in symmetric vocal folds," J. Acoust. Soc. Am. **110**, 2120–2128.

Lipshitz, H., Etheredge, R., and Glimcher, M. (**1976**). "Changes in the hexosamine content and swelling ratio of articular cartilage as functions of depth from the surface," J. Bone Joint Surg. Am. **58**, 1149–1153.

McCutcheon, S. C., Martin, J. L., and Barnwell, T. O. (**1993**). "Water quality," in *Handbook of Hydrology*, edited by D. R. Maidment (McGraw-Hill, New York), p. 11.3.

Mow, V. C., Holmes, M. H., and Lai, W. M. (**1984**). "Fluid transport and mechanical properties of articular cartilage: A review," J. Biomech. **17**, 377–394.

Mow, V. C., Kuei, S. C., Lai, W. M., and Armstrong, C. G. (**1980**). "Biphasic creep and stress relaxation of articular cartilage in compression: Theory and experiments," J. Biomed. Eng. **102**, 73–84.

Perlman, A. L., Titze, I. R., and Cooper, D. S. (**1984**). "Elasticity of canine vocal fold tissue," J. Speech Hear. Res. **27**, 212–219.

Steen, R. G., Hamer, R. M., and Lieberman, J. A. (**2007**). "Measuring brain volume by MR imaging: Impact of measurement precision and natural variation on sample size requirements," AJNR Am. J. Neuroradiol. **28**, 1119–1125.

Titze, I. R., and Hunter, E. J. (**2004**). "Normal vibration frequencies of the vocal ligament," J. Acoust. Soc. Am. **115**, 2264–2269.

Verdolini, K., Min, Y., Titze, I. R., Lemke, J., Brown, K., Mersbergen, M., Jiang, J. J., and Fisher, K. (**2002**). "Biological mechanisms underlying voice changes due to dehydration," J. Speech Hear. Res. **45**, 268–281.

Verdolini, K., Titze, I. R., and Fennell, A. (**1994**). "Dependence of phonatory effort on hydration level," J. Speech Hear. Res. **37**, 1001–1007.

Zhang, Y., Czerwonka, L., Tao, C., and Jiang, J. J. (**2008**). "A biphasic theory for the viscoelastic behaviors of vocal fold lamina propria in stress relaxation," J. Acoust. Soc. Am. **123**, 1627–1636.

An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan

Daniel Recasens[a] and Aina Espinosa
Departament de Filologia Catalana, Universitat Autònoma de Barcelona, Bellaterra, Barcelona 08001, Spain and Laboratori de Fonètica, Institut d'Estudis Catalans, c/Carme 47, Barcelona 08001, Spain

(Received 31 May 2008; revised 3 February 2009; accepted 4 February 2009)

Lingual movement data for Catalan vowel-consonant-vowel sequences reveal differences in contextual coarticulatory variability in tongue position at the middle of the consonant for /p/ > /n/ > dark /l/ > /s/ > /ʃ/ > /ɲ/ and at vowel midpoint for /u/ > /a/ > /i/. The velar stop /k/ exhibits a high degree of contextual variability in the horizontal dimension but not in the vertical dimension. These differences in coarticulatory sensitivity are attributed to differences in articulatory constraint, e.g., palatality and frication cause a higher degree of resistance in the consonant than laterality. A higher degree of contextual variability for dark /l/ than expected appears to be associated with speaker-dependent differences in darkness degree. Contextual variability is greater at regions not involved in closure or constriction formation, e.g., at the tongue dorsum than at the tongue front for alveolars. Coarticulatory resistance and coarticulatory aggressiveness are positively correlated: Phonetic segments, which are especially resistant to coarticulatory effects from the adjacent segments, exert maximal coarticulation on them. Consequently, highly constrained segments such as alveolopalatal consonants turn out to affect tongue position for less constrained segments such as back vowels rather than vice versa.
 [DOI: 10.1121/1.3089222]

PACS number(s): 43.70.Bk, 43.70.Kv [BHS] Pages: 2288–2298

I. INTRODUCTION

The concept of coarticulatory resistance in phonetic segments, i.e., the extent to which a phonetic segment blocks the coarticulatory influence of adjacent phonetic segments, has long attracted scholars in the field of speech production (Bladon and Al-Bamerni, 1976; Hardcastle and Hewlett, 1999). Coarticulatory resistance for a given consonant or vowel is a measure of its degree of articulatory variability as a function of phonetic context. According to the degree of articulatory constraint (DAC) model of coarticulation (Recasens *et al.*, 1997), coarticulatory resistance should increase with the degree of articulatory constraint, i.e., with the mechanico-inertial properties of the articulators and their involvement in the formation of a closure or constriction and with manner of articulation demands. Thus, for example, consonants produced at the alveolopalatal or palatal zone (e.g., /ɲ/) are expected to resist the coarticulatory influence of the adjacent vowels at the primary tongue dorsum (TD) articulator to a greater extent than alveolars (e.g., /n/) for the production of which this lingual region is not involved directly in closure formation. Moreover, the higher the degree of articulatory constraint, the greater the degree of coarticulatory aggressiveness, i.e., the strength of the coarticulatory effects exerted by the target segment on the neighboring phonetic segments (Farnetani, 1990; Fowler and Saltzman, 1993).

In view of literature accounts summarized below, this paper attempts to validate differences in coarticulatory resistance and aggressiveness for consonants and vowels with lingual movement data collected by means of electromagnetic midsagittal articulometry (EMA). Data will be reported for symmetrical vowel-consonant-vowel (VCV) sequences with the Catalan consonants /p, n, l, s, ʃ, ɲ, k/ and the vowels /i, a, u/. These seven consonants differ in place of articulation and in primary articulator: /p/ is bilabial; /n, l, s/ are alveolar and articulated with the tongue tip and/or the tongue blade; /ʃ, ɲ/ are usually produced at the alveolopalatal zone with the tongue blade and the tongue predorsum; and /k/ is dorsopalatovelar or dorsovelar depending on vowel context.

A. Coarticulatory resistance

1. Consonant- and vowel-dependent characteristics

The DAC model makes the following predictions regarding coarticulatory resistance for consonants and vowels.

(a) Lingual coarticulatory resistance is minimal for labials since the tongue does not intervene in the production of these consonants.
(b) Alveolopalatal stops produced with the tongue blade and dorsum (e.g., /ɲ/) are more constrained than alveolars articulated with the tip and/or blade (e.g., /n/). This ought to be so since the production of the former consonants involves a larger contact surface and a more sluggish articulator, i.e., the tongue dorsum, than the produc-

[a]Electronic mail: daniel.recasens@uab.es

tion of the latter, and fronting and raising the tongue dorsum blocks the coarticulatory activity of other tongue regions (Recasens, 1985).

(c) Manner requirements on the formation of a narrow central groove for the passage of airflow account for why lingual fricatives are highly constrained, and perhaps more so the alveolopalatal fricative /ʃ/ than the alveolar fricative /s/ for reasons mentioned in (b) (Tabain, 2001). The formation of lateral openings for the apicoalveolar lateral /l/ requires that the tongue predorsum occupies a relatively low position, thus rendering the tongue body configuration relatively constrained. Moreover, dark /l/ (which is the variety of /l/ typically found in Catalan) should be more constrained than clear /l/ since it involves additional predorsum lowering and postdorsum retraction (Recasens and Espinosa, 2005).

(d) Dorsovelars exhibit highly systematic differences in closure fronting as a function of front vs back vowels. In this case, vowel-consonant overlap yields a blended outcome showing distant closure location targets depending on the following vowel rather than a single place which adapts to vowel context to different degrees (Browman and Goldstein, 1989).

(e) The entire tongue body is highly constrained during the production of the palatal vowel /i/, much in the same way as for alveolopalatal consonants. Regarding back vowels, there may be a stronger linkage between the tongue predorsum and the primary articulator for lower pharyngeal /a/ (tongue root) than for velar /u/ (tongue postdorsum).

Vowel-dependent differences in degree of lingual coarticulation as a function of the contextual consonants are consistent with differences in articulatory constraint mentioned in (e), i.e., coarticulatory resistance varies for /i/ > /a/ > /u/ (Catalan: Recasens, 1985; German: Hoole and Kühnert, 1995; Scottish English: Zharkova, 2007). Moreover, in agreement with the consonant-dependent characteristics in degree of articulatory constraint indicated in (a)–(d), F2 data and tongue dorsum contact data for Catalan consonants obtained by means of electropalatography (EPG) reveal the existence of differences in degree of vowel coarticulation varying in the progression /p/ (minimal degree of articulatory constraint or DAC value)>/n/ (intermediate DAC value) >dark /l/, /s/, /ʃ/, and /ɲ/ (maximal DAC value), with /k/ showing more or less variability depending on fixed vowel context (Recasens *et al.*, 1997). Similar patterns of coarticulatory resistance to vowel-dependent effects have been reported for other languages. According to tongue front and back movement data for German consonants, vowel coarticulation is maximal for the labials /p, b, m, f, v/, minimal for the palatoalveolar /ʃ/, and intermediate for alveolars in the progression /d/, /n/, clear /l/ > /t/ > /s/, and for the velars /k, g/ (Hoole *et al.*, 1990). EMA data for American English /b, v, ð, d, z, ʒ, g/, show more variability in tongue body position for labials than for dentals and alveolars, least variability for palatoalveolars, and moderate or little variability for velars occurring mostly along the horizontal dimension (Fowler and Brancazio, 2000). Tongue body coarticulation for a subset of Scottish English consonants recorded with ultrasound indicate that coarticulatory variability is maximal for /p, f/, minimal for /k/, and intermediate for the alveolars /t, l, s, r/ with the rhotic being more variable than /t, l, s/ (Zharkova, 2007). Finally, F2 data for stops in Australian languages show differences in coarticulatory resistance for /c/ > /t/ > /p/, with /k/ exhibiting large degrees of vowel coarticulation (Butcher and Tabain, 2004). This cross-language scenario suggests that most consonants are specified for nearly invariant degrees of coarticulatory resistance (Fowler and Brancazio, 2000), and that this articulatory characteristic may account for spectral regularities across places of articulation such as locus equation slopes.

In the light of this literature survey, the present study will try to substantiate some relevant findings on coarticulatory resistance reported earlier and will address several new issues on the subject. Recasens *et al.* (1997) did not report any obvious differences in coarticulatory sensitivity among dark /l/, /s/, /ʃ/, and /ɲ/ presumably since the EPG technique provides little information about tongue dorsum position for consonants such as dark /l/ and /s/, which are produced with more or less tongue lowering. The research goal is then to investigate by means of EMA whether differences in manner of articulation for highly constrained consonants sharing the same place of articulation (i.e., between /s/ and dark /l/ and between /ʃ/ and /ɲ/) affect degree of coarticulatory resistance. Special attention will be paid to variability patterns in articulatory displacement for /ɲ/, which have not been explored in detail so far; indeed, studies that have used EMA data for uncovering information on coarticulatory resistance do not have the alveolopalatal nasal in their consonant inventories (English: Fowler and Brancazio, 2000; German; Hoole *et al.*, 1990).

Data reported in literature summarized above indicate that velars may be more or less sensitive to vowel coarticulatory effects at the tongue dorsum and, therefore, could be specified for higher or lower degree of articulatory constraint. Another research topic of the present study is to investigate tongue dorsum contextual variability for velars and whether coarticulation for these consonants occurs along the horizontal rather than the vertical dimension (Fowler and Brancazio, 2000).

2. Tongue regions and articulatory dimensions

While differences in coarticulatory resistance among vowels and consonants have been much explored, articulator-dependent differences in coarticulation have been neglected to a large extent. Regarding vowels, earlier work indicates more sensitivity at tongue regions not involved directly in constriction formation, namely, at the postdorsum than at the blade and mediodorsum for /i/ and at more anterior tongue regions than at the tongue back for /a, u/ (Perkell and Nelson, 1985; Kiritani *et al.*, 1977). As for consonants, Hoole *et al.* (1990) found vowel coarticulation for consonants to increase at lingual regions not involved in closure or constriction formation, i.e., at the tongue back for alveolars and at the tongue front for velars.

Information on the specific articulatory dimensions along which coarticulatory effects occur has been gathered

with EPG and EMA for vowels, but mostly with linguopalatal contact data, less so with articulatory movement data, in the case of consonants. As for vowels, coarticulatory effects on /i/ are mostly associated with consonants produced with a low tongue dorsum position, i.e., dark /l/ and to some extent /s/. Variability at the tongue dorsum surface occurs mostly vertically for /a/ presumably since this pharyngeal vowel allows for little room for backward tongue body movement, and antero-posteriorly for /u/, which has been attributed to contraction of the posterior genioglossus muscle (Perkell, 1990). Coarticulatory effects involve tongue front raising and stretching as a function of dentoalveolars and alveolopalatals for /a/ and /u/, and tongue dorsum raising as a function of velars in the case of the vowel /a/. Regarding consonants (Recasens, 1999), tongue dorsum contact effects occur mostly in vowel height and fronting (/i/ > /u/ > /a/) and, less so, in vowel fronting (/i/ > /a,u/) for dentals and alveolars, and, when available, in vowel height (/i,u/ > /a/) for (alveolo)palatals. On the other hand, tongue dorsum fronting for /k/ decreases in the progression /i/ > /a/ > /u/, thus suggesting that place of articulation for velars adapts continuously to vowel context. Alveolar and velar consonants may also exhibit differences in tongue tip and blade fronting for /i/ > /a/ > /u/ or /i/ > /a,u/, with much backing for /u/ occurring perhaps in order to assist in lip rounding during the consonant so that the acoustic effect induced by front cavity enlargement is enhanced.

The present study will try to replicate earlier findings on lingual coarticulation at the tongue front and tongue dorsum for vowels and consonants using lingual movement data. Special attention will be paid to tongue articulator-dependent differences in coarticulation for alveolopalatals. In principle, alveolopalatals should exhibit an analogous behavior to velars though, in this particular case, the fact that maximal articulatory constraint takes place at a more anterior portion of the primary dorsal articulator is prone to cause little coarticulation at most or all lingual regions.

B. Coarticulatory aggressiveness

This investigation will also explore whether there is a direct relationship between coarticulatory resistance and coarticulatory aggressiveness, i.e., whether those sounds that are most resistant should be most aggressive and those that are least resistant should barely influence the adjacent phonetic segments. Systematic surveys of this relationship have not been carried out so far with articulatory movement data. Regarding V-to-C coarticulation, the expected trend is for coarticulatory aggressiveness to be greater for /i/ than for /a, u/. On the other hand, C-to-V coarticulatory effects ought to increase with the degree of coarticulatory resistance and, therefore, with the degree of articulatory constraint for the consonant in the progression alveolopalatals > alveolars > labials, with consonants of different manners of articulation within each place category exerting different degrees of coarticulation as well.

The coarticulatory outcome is especially relevant in strings of consonants and vowels involving antagonistic articulatory configurations. In sequences of phonetic segments specified for different degree of articulatory constraint, highly constrained consonants involving a high and front tongue body position (alveolopalatals) should affect vowels leaving a low tongue front relatively free to coarticulate (/a, u/), while the opposite, i.e., tongue dorsum lowering from back vowels on alveolopalatal consonants, ought to be less prone to occur. In sequences composed of highly constrained segments, stricter requirements on consonants than on vowels to achieve a closure or constriction ought to result into more prominent C-to-V than V-to-C effects. Therefore, the prediction is for consonants that are articulated with a relatively low tongue dorsum position such as dark /l/ and /s/ to exert tongue dorsum lowering effects on /i/, rather than /i/ to exert tongue dorsum raising effects on dark /l/ and /s/.

II. METHOD

Three male speakers of Eastern Catalan of 40–50 years of age (D. Recasens, J. Pi, and J. Cererols) read symmetrical VCV combinations with the consonants /p, n, l, s, ʃ, ɲ, k/ and the vowels /i, a, u/. As pointed out in the Introduction, /l/ is generally dark, and /ɲ/ is alveolopalatal, in this Catalan dialect. All VCV sequences were embedded in the meaningful sentence [ˈgraβə pVCVp əˈβans] ("He records pVCVp earlier") and, therefore, were preceded and followed by the phonetic segments [p] and [ə], which are largely unspecified for tongue position. Speakers were instructed to pronounce all target [pVCVp] sequences ten times each with a similar degree of prominence on the first and second syllables. The same or a similar speech material produced by the same speakers was used in previous EMA and EPG investigations (Recasens, 2002; Recasens *et al.*, 1997).

Articulatory movement data were collected by means of a Carstens articulograph system AG-100. The system is composed of a head mount with three magnetic transmitter coils and several small transducer coils attached to the articulatory structures. In the present experiment, coils were placed at roughly equidistant positions on the tongue tip (TT), tongue blade (TL), and tongue dorsum (TD), as well as on the lower incisors (jaw) and on the upper lip and lower lip. The TT coil was placed about 1 cm from the apex of the tongue, and cross-speaker distance ranges between coils were 1.5–2.3 cm (TL-TT) and 1.4–1.9 cm (TD-TL). Palate traces were recorded for each speaker. Coils were also attached to the bridge of the nose and upper incisors for head movement correction. Figure 1 presents sensor locations relative to the hard palate for each speaker.

Movement and acoustic data were digitized at a sampling rate of 250 Hz for movement and at 10 kHz for speech. The kinematic data were corrected for head movement, rotated to the corresponding occlusal plane, extracted into *X* and *Y* articulatory channels, and smoothed by a finite impulse response low pass filter using a Kaiser window with a cut-off frequency of 25 Hz.

Segmentation of the VCV sequences involved identifying five temporal points on the acoustic waveform with the help of spectrographic displays, namely, V1 onset, C onset, closure offset (just for the two stops /p/ and /k/), V2 onset, and V2 offset. V1 onset was taken to occur at the onset of the

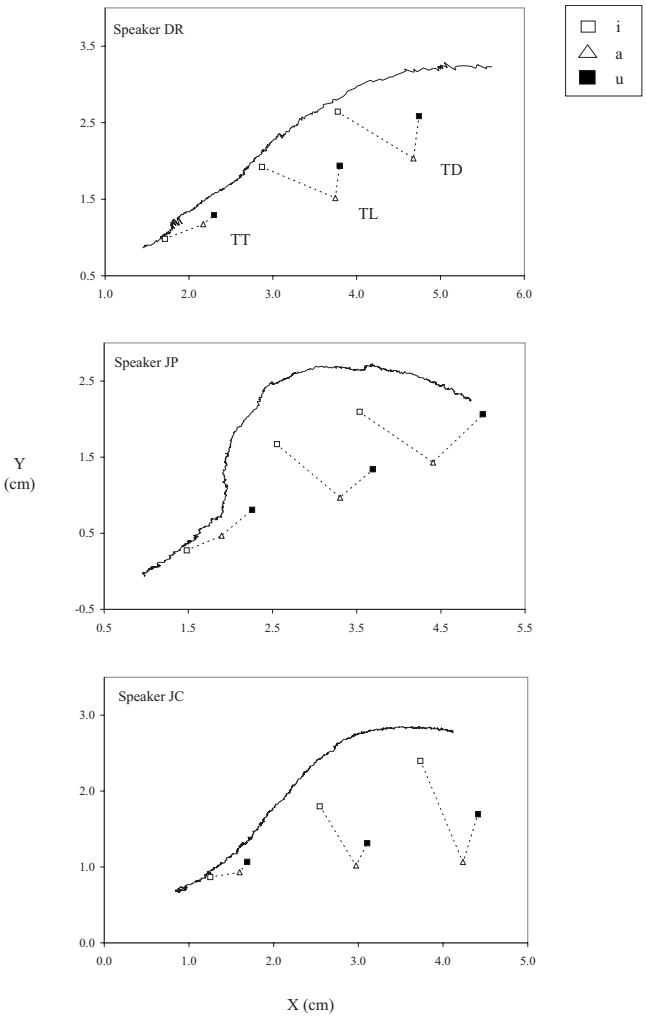


FIG. 1. TT, TL, and TD positions at consonant midpoint for /ini, ana, unu/ with palate trace. Data correspond to the speakers: D. Recasens (DR), J. Pi (JP), and J. Cererols (JC).

first pitch pulse following the burst for the first /p/, and V2 acoustic offset at the offset of vowel formant structure before closure onset for the second /p/. The onset of the intervocalic consonant occurred at the end of the V1 formant structure, which coincides with the onset of the acoustic closure for /p, k/, the frication noise for /s, ʃ/, and the low intensity formants for /n, l, ɲ/. Closure offset for /p, k/ takes place at the onset of the stop burst frication noise. V2 onset was set at the first vowel pitch pulse after the burst of a stop, the frication noise of a fricative, and the low intensity formant period of a nasal or a lateral.

Measurement points were selected according to an acoustic criterion rather than to an articulatory one since articulatory displacement maxima for the consonant and the two vowels were hard or impossible to detect in VCV sequences involving little movement (e.g., /iʃi/ and /iɲi/). Coarticulation was analyzed at two temporal points, i.e., at consonant midpoint for measuring V-to-C coarticulation and at vowel midpoint for measuring C-to-V coarticulation. Consonant midpoint was taken to occur at the midpoint of the period between C onset and closure offset for the stops /p, k/ and between C onset and V2 onset for the other consonants. Vowel midpoint was halfway between the onset and offset of voicing for the vowel. In order to render the vowel measures comparable across sequences with consonants of different manners of articulation, the 10–40 ms stop burst for unaspirated stops were not considered part of the vowel (see also Peterson and Lehiste, 1960 and Stevens, 1998, p. 258). Movement data for consonants and vowels were processed separately for each pellet and for the *X* and *Y* dimensions of articulatory displacement at the temporal points just mentioned.

A measure of coarticulatory resistance was calculated for each consonant and vowel, and for each speaker. Coarticulatory resistance for consonants was obtained at C midpoint by computing the centroid or grand mean across all three vowel contexts, and then averaging the Euclidean distances between the (x,y) position values for each vowel context and the centroid. Coarticulatory resistance for vowels was obtained at the midpoints of V1 and V2 by calculating the centroid or grand mean across all seven consonant contexts, and averaging the Euclidean distances between the (x,y) position values for each consonant context and the centroid. In both cases, the higher the degree of coarticulatory resistance, the smaller the contextual dispersion from the centroid. Data on coarticulatory resistance for vowels will be provided mostly for V1 since differences in coarticulatory resistance among vowels were found to be better defined at V1 midpoint than at V2 midpoint.

In order to find out whether variability was greater along the *X* or *Y* dimension, we also calculated the size of the contextual ranges in *X* and *Y* articulatory positions for each consonant and vowel, and for each speaker. Contextual ranges for each consonant were calculated at C midpoint by subtracting the minimal from the maximal mean contextual values across tokens for the three sequences /iCi/, /aCa/, and /uCu/ (e.g., /ipi, apa, upu/ for /p/). Ranges for each vowel at V1 location were derived by applying the same procedure to all seven VCV sequences exhibiting the same vowel (e.g., /ipi, ini, ili, isi, iʃi, iɲi, iki/ for /i/). Two other variability measures, i.e., standard deviation and coefficient of variation, were also computed but will not be reported because they yielded essentially the same differences among consonants and vowels as the contextual ranges.

A measure of coarticulatory aggressiveness was calculated for each consonant at V1 midpoint, and for each vowel at C midpoint, on the data for each speaker using the same Euclidean distance method described above. Coarticulatory aggressiveness for consonants was determined by averaging the Euclidean distances between the (x,y) position values for each consonant in the three vowel contexts and the centroid. Coarticulatory aggressiveness for vowels was computed by averaging the Euclidean distances between the (x,y) position values for each vowel in the seven consonant contexts and the centroid. In both cases, the higher the degree of coarticulatory aggressiveness, the smaller the contextual dispersion from the centroid.

Differences in coarticulatory resistance and aggressiveness were evaluated statistically by running analyses of variance (ANOVAs) with repeated measures on the Euclidean distances with consonant as a factor (resistance and aggressiveness for consonants) and with vowel as a factor (resis-

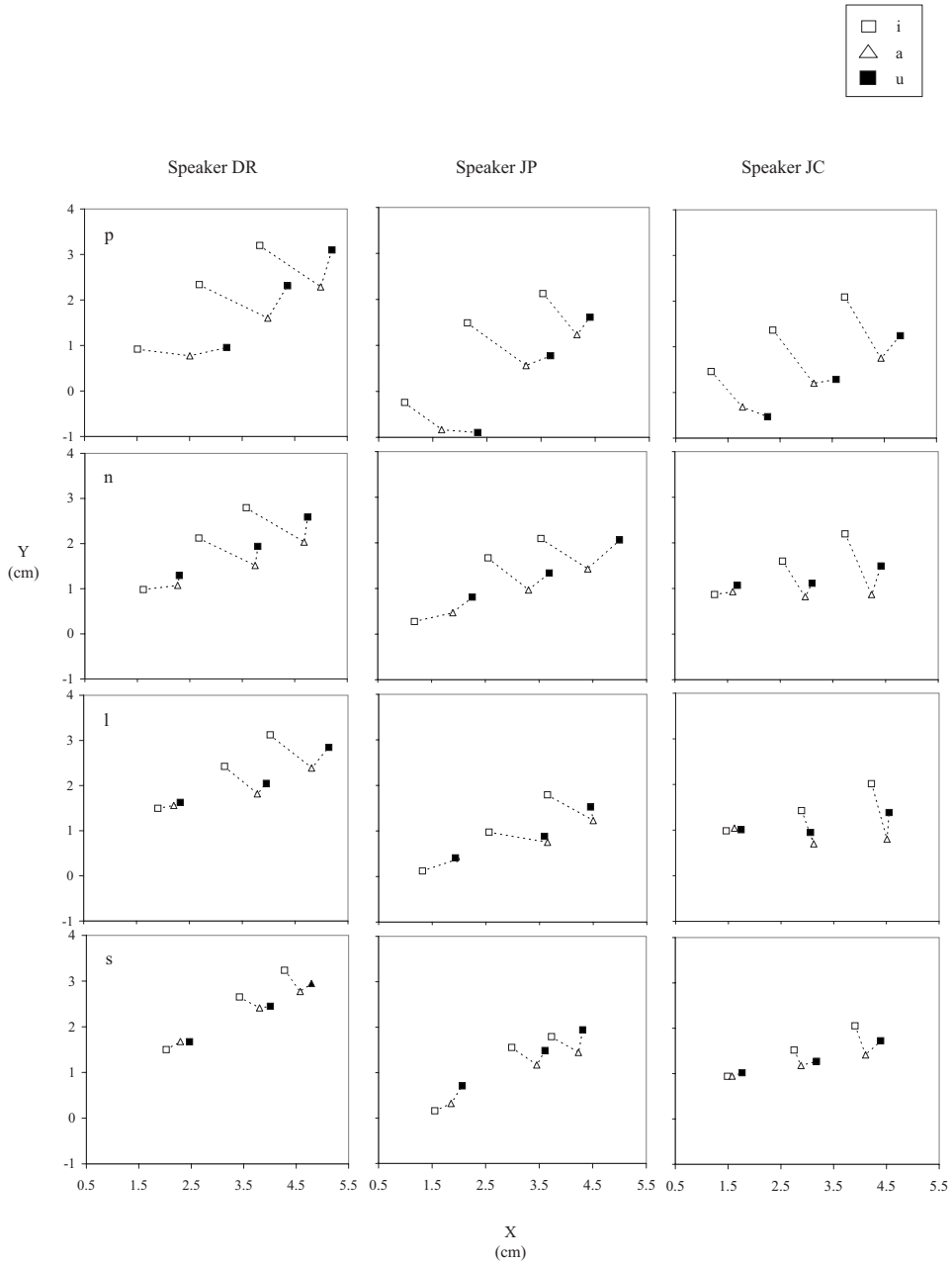


FIG. 2. (a) TT, TL, and TD positions for /p, n, l, s/ measured at consonant midpoint in the vowel contexts /i, a, u/. Data are given for all three speakers: D. Recasens (DR), J. Pi (JP), and J. Cererols (JC).

tance and aggressiveness for vowels). In order to elicit statistical differences in the V-to-C and C-to-V coarticulatory effects in articulatory position, ANOVAs with repeated measures were run separately on the *X* and *Y* lingual position values with vowel as a factor (V-to-C effects) and with consonant as a factor (C-to-V effects). Results for the significant C-to-V effects will be provided both at V1 and at V2 so as to achieve a more accurate description of the extent to which a given vowel is affected by different consonants. Levels of the vowel factor were i, a, and u and levels of the consonant factor were p, n, l, s, ʃ, ɲ, and k. Each speaker contributed one averaged score per condition. Huynh–Feldt corrected degrees of freedom were performed on the main effects. Given the small number of data entered in the ANOVAs, pairwise comparisons between levels of a given factor were carried out using Fischer least significant difference *post-hoc* tests. The degree of significance was set at $p < 0.05$.

III. RESULTS

A. Coarticulatory resistance

1. Consonants

According to Fig. 2, tongue tip position at closure or constriction location is highest for the alveolars /n, l, s/, and somewhat more retracted for /s/ than for the two other consonants; tongue blade and tongue dorsum position turns out to be higher for /s/ than for /n, l/. Bilabials, alveolopalatals, and velars are produced with the tongue tip down. Place of articulation for /ʃ/ and /ɲ/ occurs at the tongue blade and perhaps at the tongue dorsum, as indicated by the fact that

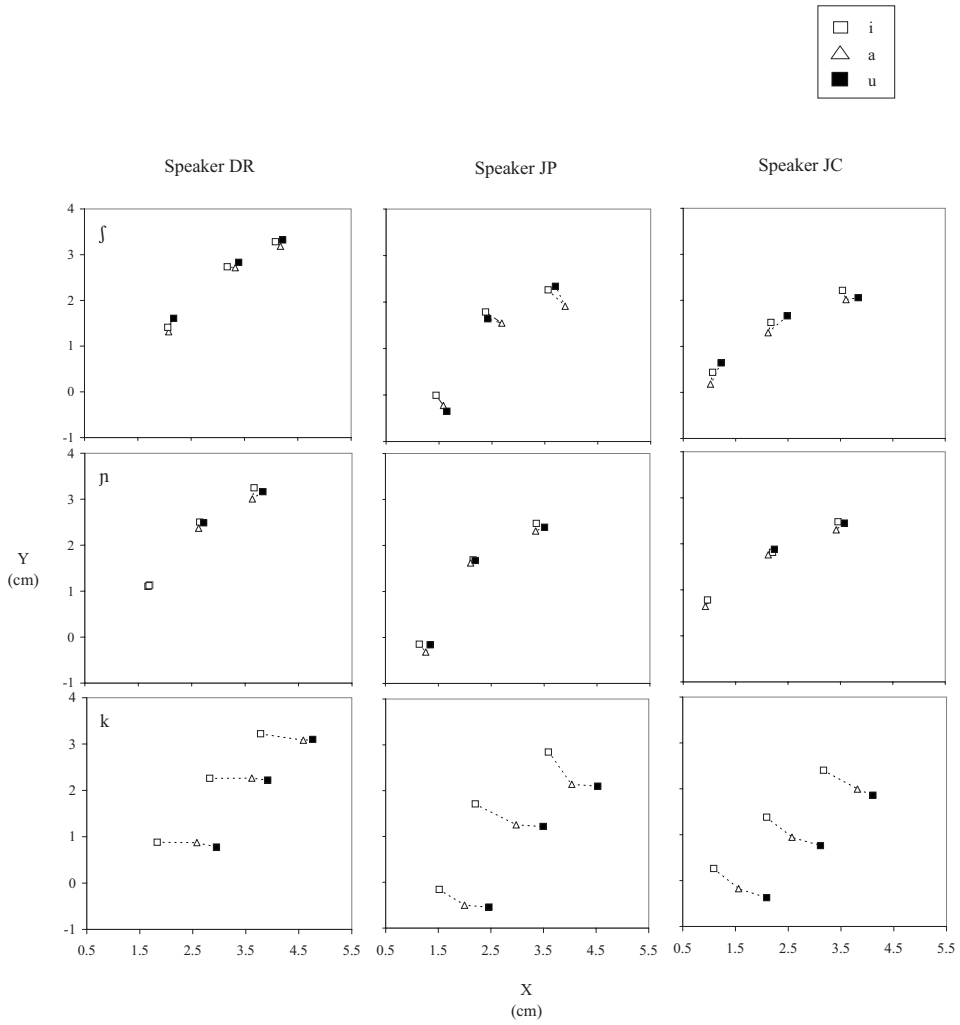


FIG. 2. (b) TT, TL, and TD positions for /ʃ, ɲ, k/ measured at consonant midpoint in the vowel contexts /i, a, u/. Data are given for all three speakers: D. Recasens (DR), J. Pi (JP), and J. Cererols (JC).

both articulators occupy the highest and anteriormost position in this case. The highest tongue position for /k/ is at the dorsal articulator.

ANOVAs on mean Euclidean distances yielded a significant effect of consonant for all articulators TT [$F(2.9,5.9)=21.39$, $p=0.001$], TL [$F(6,12)=46.00$, $p=0.000$], and TD [$F(1.9,3.9)=30.09$, $p=0.004$]. Consonant-dependent differences in coarticulatory sensitivity presented in Fig. 3 (top) turned out to be in accordance with data for Catalan reported in earlier EPG and acoustic studies. According to results from *post-hoc* tests, those differences vary in the progression /p/ > /n/ > /l,k/ > /s/ > /ʃ,ɲ/ especially for the TL and/or TD movement data. These consonant-dependent differences in coarticulatory sensitivity become apparent when the contextual position values for each consonant in Fig. 2 are taken into consideration: /p/ is most variable, /ʃ, ɲ/ are least variable, and, among alveolars, variability is maximal for /n/ and minimal for /s/. Considerable variability for /l/ appears to be related to speaker-dependent differences in tongue blade and dorsum positions mostly in the context of /i/ and, presumably, in darkness degree: Thus, the tongue body for the alveolar lateral occupies a higher position next to /i/ than to /a, u/ for some speakers (D. Recasens and J. Cererols) but less so or not at all for others (J. Pi); therefore, the consonant is less resistant to tongue dorsum raising effects induced by the high front vowel and thus, presumably clearer, for the two former speakers than for the latter. Speaker-dependent contextual differences for /l/ also occur along the *X* dimension.

Figure 3 (bottom) shows that the consonant-dependent differences in coarticulatory variability just referred to hold both along the *X* and *Y* dimensions. The only clear exception is /k/, which favors more *X* than *Y* movement and, therefore, is highly variable along the *X* dimension (as for /n/) and highly resistant along the *Y* dimension (as for /s/) (see also Fig. 2). The labial /p/ and the alveolars /n, l, s/ also exhibit a slightly greater degree of variability in the horizontal dimension than in the vertical dimension at the tongue tip and blade. Both dimensions of articulatory displacement show similar degrees of variability at all lingual regions for the alveolopalatals /ʃ, ɲ/.

Figure 3 (bottom) also reveals a decrease in contextual variability for TD > TL > TT in the case of the consonants /p, n, l, s/ along the *Y* dimension, and of /n, l/ and, to some extent, /s/ along the *X* dimension. It thus appears that the degree of coarticulation for front lingual consonants increases the more we depart from the primary articulator. The

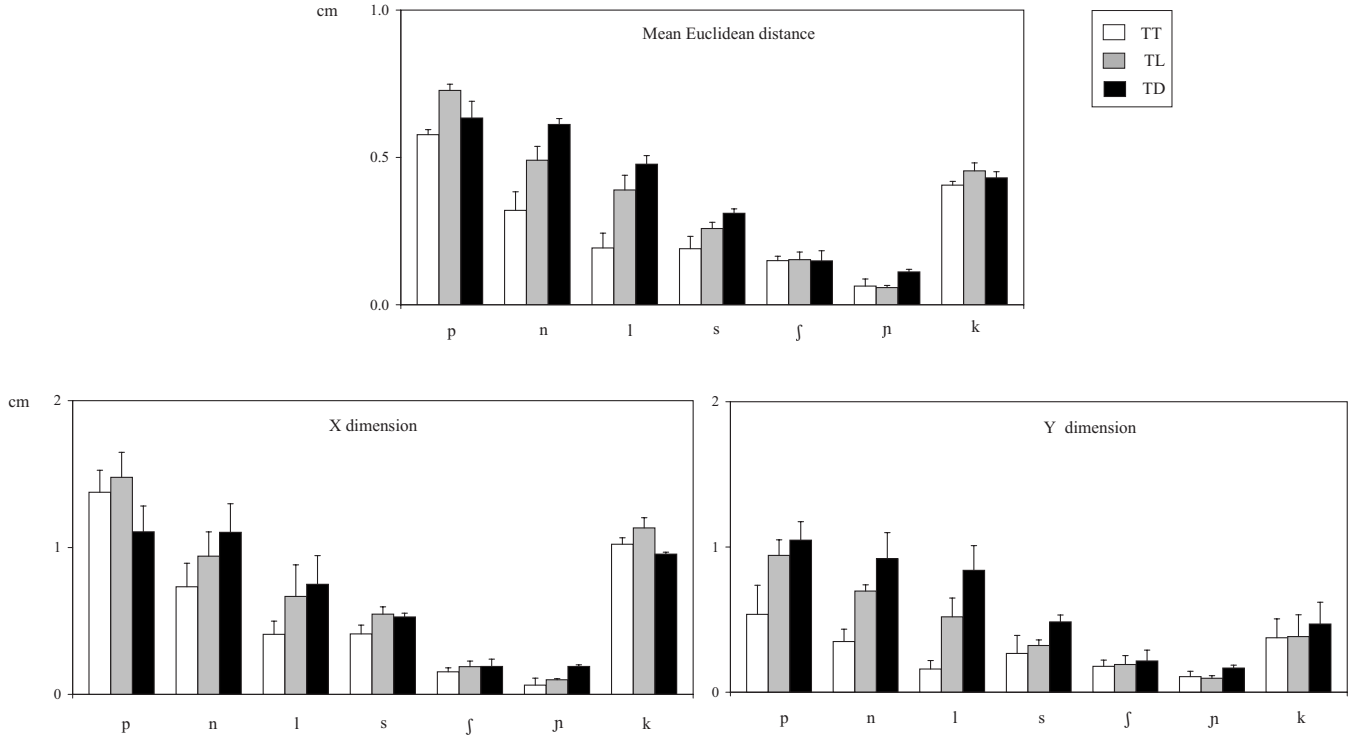


FIG. 3. Mean Euclidean distances in TT, TL, and TD positions (top), and ranges of tongue position values along the *X* and *Y* dimensions (bottom), for /p, n, l, s, ʃ, ɲ, k/ across vowel contexts measured at consonant midpoint. Data correspond to means and standard deviations across speakers.

lamino-predorsal and dorsal consonants /ʃ, ɲ, k/ exhibit no clear variability differences among lingual articulators (though they may also show a trend for the *X* and *Y* ranges for the tongue dorsum to exceed those for the tongue tip and blade).

According to Table I (see also Fig. 2), labials (/p/) and velars (/k/) allow significant vowel effects in horizontal displacement for /i/ > /a/ > /u/ at all or most pellets, which in the case of /k/ could be associated with three instead of two places of articulation. Moreover, labials also show significant tongue dorsum effects in vertical displacement as a function of high versus low vowels. *X* and *Y* vowel effects on the alveolopalatal consonants /ʃ, ɲ/ are negligible. As for alveolars, tongue fronting decreases in the progression /i/ > /a, u/ (/n/) and /i, a/ > /u/ (/s/), and tongue dorsum height in the progression /i, u/ > /a/ (/n, l/) and /i/ > /a/ (/s/). In a few cases, vowel-dependent differences in articulatory position for a given alveolar consonant fail to achieve significance because they may show up to different degrees or fail to occur depending on speaker, e.g., /i/ > /a, u/ for /p/ (TT*Y* and TL*Y*) and for /l/ (TL*Y*) (see Fig. 2).

TABLE I. Significant main vowel-dependent effects and *post-hoc* pairwise comparisons for /p, n, l, s, ʃ, ɲ, k/ as a function of tongue articulator and articulatory dimension (*, $p<0.05$; **, $p<0.01$; ***, $p<0.001$). *X* differences among contextual vowels proceed from most anterior to most posterior, and *Y* differences from highest to lowest.

	TT*X*	TL*X*	TD*X*	TT*Y*	TL*Y*	TD*Y*
	**	**	**			**
p	i > a > u	i > a > u	i > a > u			i, u > a
	*	**	**		*	*
n	i > a, u	i > a, u	i > a, u		i, u > a	i, u > a
	*		*			*
l						i, u > a
	*	*	*		*	*
s	i, a > u	i, a > u	i > u		i > a	i > a
	*					
ʃ	i > u					
		**			*	
ɲ		i, a > u			u > a	
	**	***	*			
k	i > a > u	i > a > u	i > a, u			

2. Vowels

Tongue tip, blade, and dorsum position at V1 midpoint conform to the expected differences in anteriority for /i/ > /a/ > /u/ and in height for /i/ > /u/ > /a/. Thus, as shown in Fig. 4, the tongue dorsum is located between 3.5 and 4 cm behind the origin of the coordinate space at the upper incisors for /i/, between 4 and 4.5 cm for /a/, and between 4 and 5 cm for /u/. Moreover, according to the same figure, tongue dorsum height is about 2.5 cm for /i/, between 2 and 2.5 cm for /u/, and between 1.5 and 2 cm for /a/.

According to statistical tests run at V1 midpoint, differences in degree of consonant-dependent variability for vowels were highly significant at TT [$F(1.1, 2.3)=34.03$, $p=0.021$] and TL [$F(1, 2)=39.55$, $p=0.024$], and barely significant at TD [$F(1, 2)=13.42$, $p=0.067$]. *Post-hoc* tests reveal that those differences decrease in the progression /u/ > /a/ > /i/ at the tongue tip and blade and /u/ > /i/ at the tongue dorsum (see Fig. 5, top). Figure 4 shows indeed much less dispersion among consonant contexts for the high front vowel than for the two back vowels. All articulators exhibit similar degrees of contextual variability along the *X* and *Y*

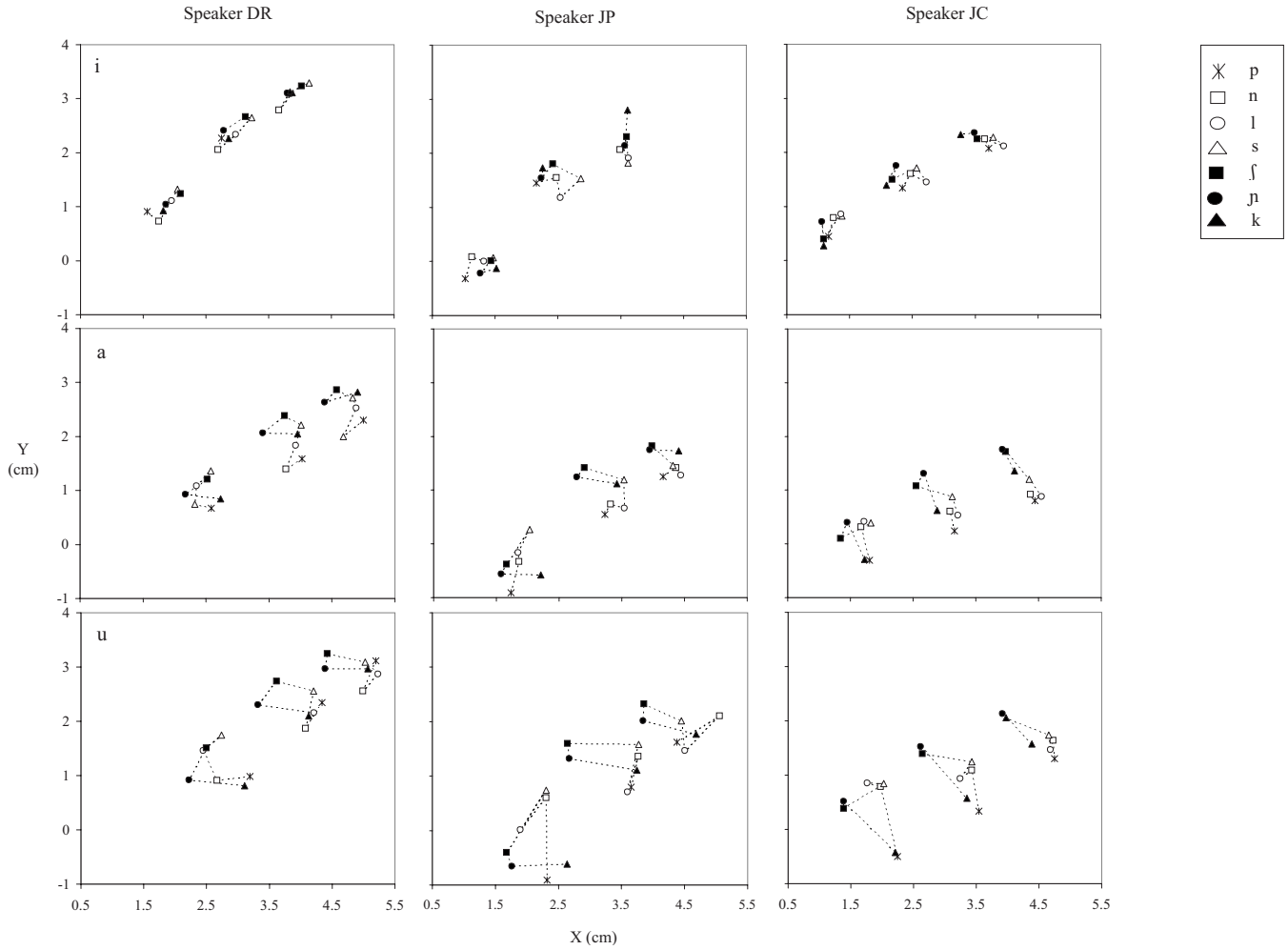


FIG. 4. TT, TL, and TD positions for /i, a, u/ measured at V1 midpoint as a function of contextual consonants of different place of articulation. Data are given for all three speakers: D. Recasens, J. Pi, and J. Cererols.

dimensions in the case of /i/ and some more variability in the *Y* dimension than in the *X* dimension for /a/ (Fig. 5, bottom). As for /u/, coarticulatory variability is especially large at the tongue tip along the *Y* dimension and slightly greater horizontally than vertically at the tongue blade and dorsum.

Table II provides information about the significant consonant effects at the two vowels' midpoint (see also Fig. 4). As indicated in the table caption, *X* and *Y* values for a given consonant appearing outside the panel always exceed those for the consonants that are paired with it and appear inside the panel. Consequently, for a given consonant pair, the consonant outside the panel may be considered to be responsible for significant C-to-V effects in the /a, u/ condition (since these effects involve tongue body raising and fronting), and the consonant inside the panel for significant C-to-V effects in the /i/ condition (since these effects involve tongue body lowering and backing). Effects on /i/ involve mostly tongue blade retraction as a function of /l, s/ (TL*X*) and tongue blade lowering as a function of /l/ (TL*Y*). Back vowels allow more considerable consonantal effects. Effects along the horizontal dimension occur on /u/ and, less so, on /a/, and involve fronting at the tongue blade and dorsum and, less so, at the tongue tip in the context of /ʃ, ɲ/, and at the tongue tip in the context of alveolars, mostly /l/. Effects along the vertical dimension are slightly larger on /a/ than on /u/, and involve raising at the tongue tip in the context of /l/, at the tongue blade next to /ɲ/, at the tongue blade and dorsum next to /k/, and at all articulators in the context of /s, ʃ/ (more so at tongue tip and blade in the case of /s/ and at tongue blade and dorsum in the case of /ʃ/).

B. Coarticulatory aggressiveness

ANOVAs on mean Euclidean distances yielded a significant consonant effect at V1 midpoint for all articulators TT [$F(1.6,3.3)=16.24$, $p=0.020$], TL [$F(6,12)=49.76$, $p=0.000$], and TD [$F(2.8,5.6)=20.75$, $p=0.002$]. Significant differences in coarticulatory variability take place for /p/ > /n/ > /l, s/ > /ʃ, ɲ/, with /k/ not differing significantly from /p, l, s/, and therefore resemble those reported at C midpoint (see Fig. 6, top). Indeed, consonants such as /p/ and /n/ exhibit larger differences in articulatory position across vowel contexts than consonants such as /ʃ/ and /ɲ/ and, therefore, the former can be considered to be less aggressive than the latter. A correlation analysis between mean Euclidean distances for consonants at V1 midpoint and at C midpoint yielded a high $\underline{r}$ value between the two measures (0.931),

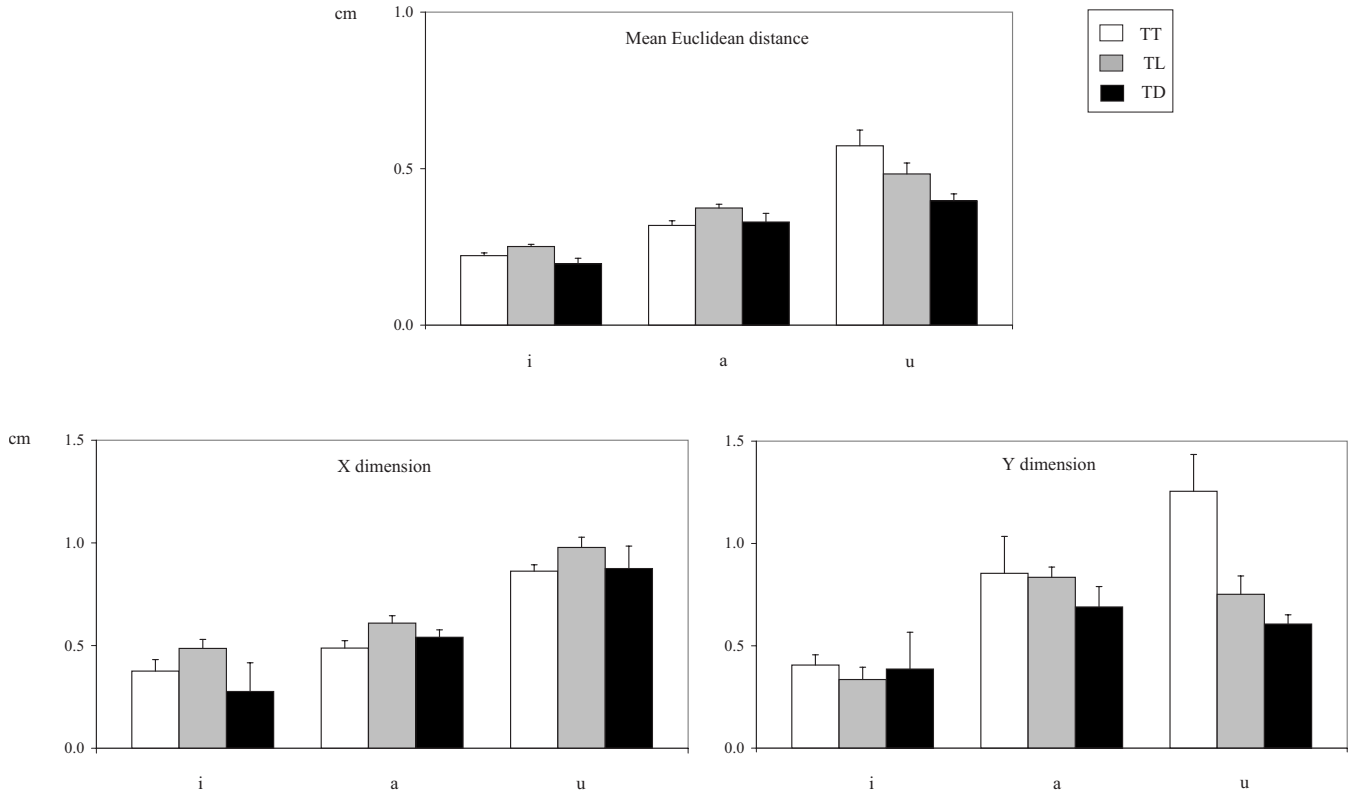


FIG. 5. Mean Euclidean distances in TT, TL, and TD positions (top), and ranges of tongue position values along the *X* and *Y* dimensions (bottom), for /i, a, u/ across consonant contexts measured at V1 midpoint. Data correspond to means and standard deviations across speakers.

thus proving that coarticulatory aggressiveness for consonants varies directly with coarticulatory resistance to vowel effects.

A similar scenario holds between coarticulatory aggressiveness and coarticulatory resistance for vowels. According to statistical tests run at C midpoint, the vowel factor turned out to be significant for TT [$F(1.8,3.6)=21.90$, $p=0.010$] and TL [$F(1.9,3.8)=38.60$, $p=0.003$], and barely significant for TD [$F(1,2)=21.48$, $p=0.043$]. *Post-hoc* tests reveal the existence of significant differences in coarticulatory variability for /a,u/ > /i/ at TL and TD (see Fig. 6, bottom). Given that /i/ exhibits less considerable consonant-dependent differences in articulatory position than /a, u/, it may be ascertained that the former vowel exerts greater V-to-C effects than the two latter ones. A correlation analysis between mean Euclidean distances for the three vowels at C midpoint and at V1 midpoint yields a high $\underline{r}$ value (0.873), meaning that coarticulatory aggressiveness and coarticulatory resistance for vowels are closely related.

The relative strength of the C-to-V and V-to-C effects in sequences of consonant and vowel segments produced with antagonistic articulatory actions depends on degree of articulatory constraint. Thus, as predicted, C-to-V effects from alveolopalatals on /a, u/ (Table II) are more prominent than V-to-C effects from /a, u/ on alveolopalatals (Table I). The coarticulation scenario in sequences of highly constrained segments conforms to some extent to our initial prediction that C-to-V effects should prevail onto V-to-C effects. Indeed, data show some tongue dorsum lowering and retraction effects from dark /l/ and /s/ on /i/ (Table II), and tongue dorsum raising and fronting effects from /i/ on /s/ but much less so on /l/ (Table I).

IV. DISCUSSION AND CONCLUSIONS

Patterns of contextual variability in tongue position for consonants and vowels (as well as individual V-to-C and C-to-V effects) reported in the present study confirm what appears to be a general principle of coarticulatory behavior, namely, that coarticulatory resistance and aggressiveness are indicative of degree of articulatory constraint and depend on the degree of involvement of a given articulatory region in closure and constriction formation and on manner requirements.

Differences in lingual variability for /p/ > /n/ > /l,k/ > /s/ > /ʃ,ɲ/ and /u/ > /a/ > /i/ are in agreement with the scale of degree of articulatory constraint proposed by the DAC model. They confirm previous accounts in that coarticulatory resistance decreases for alveolopalatals > alveolars > labials both along the horizontal and vertical directions of articulatory displacement (see below regarding the status of velars). In contrast with EPG data adduced by Recasens *et al.* (1997), a higher degree of coarticulatory resistance for alveolopalatals than for alveolars has been found to hold among the highly constrained consonants dark /l/, /s/, /ʃ/, and /ɲ/. These differences in coarticulatory sensitivity indicate that frication contributes more than laterality and possibly darkness to an increase in degree of articulatory constraint. Similar findings have been reported in studies measuring lingual coarticulatory resistance and variations in jaw-tongue tip cohesion induced by changes in vocal effort for German consonants, i.e., in this case variability was

TABLE II. Significant consonant-dependent effects and *post-hoc* comparisons for /i, a, u/ as a function of tongue articulator and articulatory dimension. (Top of each panel) Main effects at V1 and V2 (*, $p<0.05$; **; $p<0.01$; ***, $p<0.001$). (Bottom of each panel) Consonant pairs yielding a significant effect at either V1 or V2 (in regular typeface) or at both V1 and V2 (in boldface). The consonants appearing outside the panel exhibit a more anterior (*X*) or higher (*Y*) position value than the ones placed on the same row inside the panel.

		TT*X*	TL*X*	TD*X*	TT*Y*	TL*Y*	TD*Y*
	V1	NO	*	NO	NO	NO	NO
	V2	NO	*	NO	NO	NO	NO
i	p		**l**				
	n						
	l						
	s					l	
	ʃ						
	ɲ		**l s**			p l k	
	k		**s**				
	V1	**	***	***	**	**	**
	V2	NO	**	**	**	***	***
a	p						
	n	s	l				
	l	s			p		
	s				**p ʃ ɲ k**	**p l**	p l
	ʃ		s	l s	p k	**p n l s k**	**p** n
	ɲ	n l s k	**p n l** s	n **l s**		p **n**	
	k		s			**p** l	**p** l
	V1	***	***	**	**	*	*
	V2	***	***	***	**	***	*
u	p						
	n	**k**					
	l	p n s **k**	s k		ɲ k		
	s	**k**			p n l **k**	n l **k**	l
	ʃ	**p k**	**p** n l **s k**	p **n l** s k	p	**l k**	**l s k**
	ɲ	p n **s k**	**p n l s ʃ k**	**p n l s ʃ k**		l	
	k						s

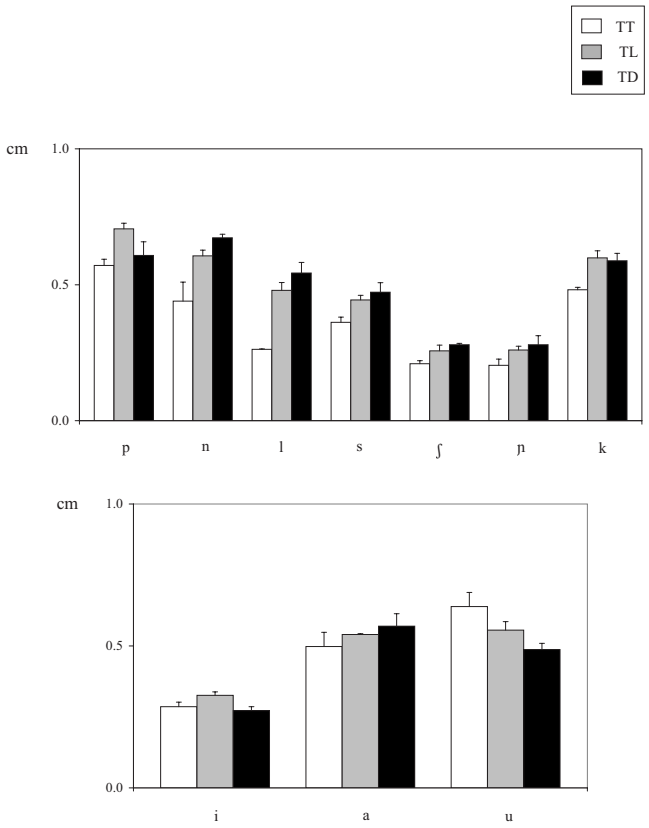


FIG. 6. Mean Euclidean distances in TT, TL, and TD positions for consonants across vowel contexts measured at V1 midpoint (top) and for vowels across consonant contexts measured at consonant midpoint (bottom). Data correspond to means and standard deviations across speakers.

found to be less for /s/ than for /l/ though the same for /l/ and /n/ perhaps since /l/ is clear in this language (Hoole *et al.*, 1990; Mooshammer *et al.*, 2006). In Scottish English, however, no differences in lingual coarticulatory resistance were found to occur between /l/ and /s/ (Zharkova, 2007).

The consonant /l/ turned out to be more variable than expected in the present study presumably in line with speaker-dependent differences in darkness degree. This appears to be in agreement with EPG and acoustic data reported in earlier studies for dark /l/ in Eastern Catalan (Recasens *et al.*, 1995) showing a high degree of speaker-dependent variability in the sequence /ili/ both regarding dorsopalatal contact size (between 10% and 35% of contact activation over the total amount of electrodes at the palatal zone) and F2 (between about 1250 and 1450 Hz). This finding suggests that certain articulatory properties are allowed to vary more than others, e.g., laterality and darkness than frication.

The special status of velars has been confirmed by the present investigation. Data from previous studies show that this consonant may be more or less variable depending on factors such as articulatory direction (Fowler and Brancazio, 2000), the transconsonantal vowel in asymmetrical VCV sequences (Recasens *et al.*, 1997), and language (Butcher and Tabain, 2004). In agreement with Fowler and Brancazio (2000), much contextual variability for velars has been found to take place along the horizontal dimension for /i/ > /a/ > /u/, which is in support of the notion that these consonants blend with the adjacent vowels. A high degree of coarticulatory resistance along the *Y* dimension is in accordance with minimal V-to-C coarticulatory effects on /k/ reported elsewhere (Zharkova, 2007).

As for vowels, fronting and raising the tongue body for /i/ appears to constrain all tongue regions to a large extent thus rendering this vowel more coarticulation resistant than back vowels. In agreement with previous studies, differences in coarticulatory resistance also hold for /a/ > /u/.

Whenever available, coarticulatory effects occur mostly at regions not participating in closure or constriction formation. Indeed, variability for /n, l, s/ was greater at the tongue dorsum where *X* and *Y* effects may take place as a function of front versus back and high versus low vowels, respectively, than at the tongue tip and blade where effects occur generally along the *X* dimension as a function of front versus back vowels. Regarding vowels, consonantal effects on /a, u/ take place mostly at the tongue tip and blade as a function of alveolopalatals and alveolars through variations in tongue height for /a/ and in tongue fronting for /u/, which is in accordance with a muscle-based account of variability for vowels (Perkell, 1990).

Trends in coarticulatory aggressiveness were found to be also consistent with the DAC model of coarticulation. Both for consonants and vowels, phonetic segments which were more resistant turned out to be more aggressive as well. Specific coarticulatory effects between segments involving antagonistic articulatory positions were in agreement with this general principle: More constrained alveolopalatals resist the lowering and backing influence of less constrained /a, u/ while exerting maximal coarticulation on these vowels. In consonant-vowel sequences composed of two highly constrained and antagonistic segments, e.g., dark /l/ or /s/ and /i/, the consonant appears to prevail onto the vowel rather than the other way around, which is in line with the articulatory demands for the two sound classes. In any case, the degree of V-to-C coarticulation exerted from /i/ on the consonant appears to be hard to evaluate since it depends heavily on the exact degree of darkness (for /l/) and frication (for /s/), which may vary according to speaker and other factors.

In summary, consonants and vowels may be differentiated quite successfully in terms of coarticulatory resistance, which falls naturally from the production demands involved. Moreover, a direct relationship appears to hold between coarticulatory resistance and coarticulatory aggressiveness. This scenario suggests that the coarticulatory resistance scale is a valid criterion of consonant and vowel classification and provides valuable information about the planning mechanisms used by speakers in speech production.

The present research points to several issues for further analysis. It remains unclear why phonetic segments in VCV sequences are more variable if produced with a low tongue dorsum than with a high dorsum. This finding suggests that production requirements for these sequences differ from those for consonant clusters where manner demands for low tongue dorsum consonants overcome consonants articulated with a high tongue dorsum position (Recasens and Pallarès, 2001). It may be that articulatory control on certain production mechanisms is greater in consonant clusters than in CV and VC sequences because consonants are required to configure the articulatory structures more precisely than vowels for the implementation of specific manners of articulation. In consonant clusters, the integrity of the consonant appears to be affected to a larger extent by deviations in tongue dorsum position induced in dark /l/, the apical trill /r/, and /s/ than in alveolopalatals even though the tongue dorsum is the primary articulator of the latter but not of the former group of consonants.

ACKNOWLEDGMENTS

This research was funded by Project No. HUM2006-03743 of the Ministry of Science and Technology of Spain and FEDER, and Project No. 2005SGR864 of the Generalitat de Catalunya. We would like to thank three anonymous reviewers and Joaquín Romero for comments.

Bladon, R. A., and Al-Bamerni, A. (**1976**). "Coarticulation resistance in English /l/," J. Phonetics **4**, 135–150.
Browman, C. P., and Goldstein, L. (**1989**). "Articulatory gestures as phonological units," Phonology **6**, 201–251.
Butcher, A., and Tabain, M. (**2004**). "On the back of the tongue: Dorsal sounds in Australian languages," Phonetica **61**, 22–52.
Farnetani, E. (**1990**). "V-to-V lingual coarticulation and its spatiotemporal domain," *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 93–130.
Fowler, C. A., and Brancazio, L. (**2000**). "Coarticulatory resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation," Lang Speech **43**, 1–41.
Fowler, C., and Saltzman, E. (**1993**). "Coordination and coarticulation in speech production," Lang Speech **36**, 171–195.
Hardcastle, W. J., and Hewlett, N. (**1999**). *Coarticulation. Theory, Data and Techniques* (Cambridge University Press, Cambridge).
Hoole, P., and Kühnert, B. (**1995**). "Patterns of lingual variability in German vowel production," *Proceedings of ICPhS95*, edited by K. Elenius and P. Branderud (KTH, Stockholm University, Stockholm), pp. 442–445.
Hoole, P., Gfroerer, S., and Tillmann, H. G. (**1990**). "Electromagnetic articulography as a tool in the study of lingual coarticulation," Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München, **28**, 107–122.
Kiritani, S., Itoh, K., Hirose, H., and Sawashima, M. (**1977**). "Coordination of the consonant and vowel articulations—X-ray microbeam study of Japanese and English," Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo **11**, 31–37.
Mooshammer, T., Hoole, P., and Geumann, A. (**2006**). "Interarticulator cohesion within coronal consonant production," J. Acoust. Soc. Am. **120**, 1–13.
Perkell, J. S. (**1990**). "Testing theories of speech production: Implications of some detailed analyses of variable articulatory data," *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 263–288.
Perkell, J. S., and Nelson, W. L. (**1985**). "Variability in production of the vowels /i/ and /a/," J. Acoust. Soc. Am. **77**, 1889–1925.
Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.
Recasens, D. (**1985**). "Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences," Lang Speech **28**, 97–114.
Recasens, D. (**1999**). "Lingual coarticulation," *Coarticulation*, edited by W. J. Hardcastle and N. Hewlett (Cambridge University Press, Cambridge), pp. 80–104.
Recasens, D. (**2002**). "An EMA study of VCV coarticulatory direction," J. Acoust. Soc. Am. **111**, 2828–2841.
Recasens, D., and Espinosa, A. (**2005**). "Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: Evidence from two Catalan dialects," J. Int. Phonetic Assoc. **35**, 1–25.
Recasens, D., and Pallarès, M. D. (**2001**). "Coarticulation, assimilation and blending in Catalan consonant clusters," J. Phonetics **29**, 273–301.
Recasens, D., Pallarès, M. D., and Fontdevila, J. (**1995**). "Co-articulatory variability and articulatory-acoustic correlations for consonants," Eur. J. Disord. Commun. **30**, 203–212.
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.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).
Tabain, M. (**2001**). "Variability in fricative production and spectra: Implications for the hyper- and hypo- and quantal theories of speech production," Lang Speech **44**, 57–94.
Zharkova, N. (**2007**). "Quantification of coarticulatory effects in several Scottish English phonemes using ultrasound," QMU Speech Science Research Centre Working Papers, No. WP13, pp. 1–19.

Attenuation of vocal responses to pitch perturbations during Mandarin speech

Hanjun Liu
Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, People's Republic of China

Yi Xu
Department of Speech, Hearing and Phonetics Sciences, Division of Psychology and Language Sciences, University College London, London WC1N 1PF, United Kingdom

Charles R. Larson[a)]
Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208

(Received 18 January 2008; revised 21 January 2009; accepted 22 January 2009)

The effect of stimulus timing on vocal responses to pitch-shifted feedback was investigated in different intonation patterns during Mandarin speech production. While speaking a four-word sentence consisting of the high-level tone, where the fundamental frequency (F_0) of the final word was either increased (question intonation) or slightly falling (statement intonation), pitch-shift stimuli ($\pm$100 cents, 200 ms duration) were presented at three different times (160, 240, or 340 ms) after vocal onset. Results showed that in the question intonation, response magnitudes (16 cents) were significantly reduced for the 340 ms condition compared to the 160 (26 cents) or 240 (23 cents) ms conditions. No significant differences were found, however, as a function of stimulus timing in the statement intonation. These findings demonstrate that a planned change in F_0 can cause a modulation in the reflexive response to a perturbation in voice pitch feedback and that there is a critical time period during which the response mechanisms are most sensitive to the planning process. These findings suggest an approach for the study of mechanisms involved in the timing of successive words during speech. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3081523]

PACS number(s): 43.70.Fq, 43.70.Bk, 43.70.Jt, 43.70.Mn [AL] Pages: 2299–2306

I. INTRODUCTION

In speech the role of voice fundamental frequency (F_0) is to convey various meanings in addition to those conveyed by consonants and vowels. Recent research on intonation has shown that F_0 can be used to encode two or more communicative meanings simultaneously (Eady and Cooper, 1986; Pierrehumbert and Hirschberg, 1990; Xu, 1999; Liu and Xu, 2005). It has been found that Mandarin speakers express lexical tone, focus, and question intonation by controlling F_0 in different ways (Xu, 1999; Liu and Xu, 2005). Tone is expressed by making F_0 approach a specific trajectory during the syllable, focus is encoded by expanding the pitch range of focused words and compressing the pitch range of post-focused words, and question is expressed by increasingly raising F_0 toward the end of the utterance. Similarly, simultaneous control of F_0 by lexical, focal, and intonational functions has been found in non-tonal languages such as English (Xu and Xu, 2005; Liu and Xu, 2007).

Important as it is to speech, little is known about neural mechanisms controlling voice F_0. Recent attempts to model this system provide a theoretical rationale on how the nervous system might accomplish this task (Hain *et al.*, 2000). An important component of such models is the role of sensory feedback in the control process. Experimental studies have demonstrated that if voice auditory feedback is missing, masked, or altered, voice F_0 control is diminished (Elliott and Niemoeller, 1970; Svirsky *et al.*, 1992; Mürbe *et al.*, 2002). The role of auditory feedback in voice control during speech development is particularly important, and deafness during language acquisition severely affects the quality of a child's articulation (Oller and Eilers, 1988). While the voice is not affected immediately if hearing loss occurs after mastering speech skills, there is a slow reduction in F_0 control (Angelocci *et al.*, 1964; Binnie *et al.*, 1982; Leder *et al.*, 1987; Waldstein, 1990; Lane *et al.*, 1997). It has also been shown that changes in kinesthetic feedback through anesthetization procedures reduce fine control of voice F_0 in pitch-masking tasks (Mallard *et al.*, 1978; Sundberg *et al.*, 1995). Recently, Larson *et al.* (2008) reported that voice responses to pitch perturbations were larger when the vocal fold mucosa was anesthetized than during normal kinesthesia, which could be due to the involvement of both kinesthesia and auditory feedback control of voice F_0. In addition, it has been demonstrated that auditory feedback plays an important role in voice control in birds and mammals. For example, presenting electronically pitch-shifted voice feedback to horseshoe bat elicits compensatory changes in call frequency (Metzner, 1996), and the Lombard effect has been demonstrated in monkeys (Sinnott *et al.*, 1975). All of these studies

[a)]Author to whom correspondence should be addressed. Electronic mail: clarson@northwestern.edu

point to the importance of auditory feedback for on-line control of speech and vocal communication in general (Smotherman, 2007).

In addition to feedback models of voice F_0 control, it has also been suggested that speech and F_0 are controlled by "internal models" (Wolpert *et al.*, 1995; Perkell *et al.*, 1997; Kawato, 1999) that represent the relation between motor commands and their acoustic output. The development and maintenance of such internal models may depend on sensory feedback (Wolpert *et al.*, 1998; Jones and Munhall, 2000). That is, the neural system learns how to generate motor commands that can produce the desired acoustic output and to continuously fine-tune the internal representation of such motor-acoustic relations. Jones and Munhall (2000, 2002) showed that both English and Mandarin speakers compensated for pitch-shifted voice feedback to maintain their habitual pitch targets when their voice auditory pitch feedback was slowly shifted up or down without their awareness, indicating that auditory feedback is used in calibrating the internal model for the control of F_0. When the pitch-shifted voice feedback was then suddenly removed, there was a negative aftereffect, which further supported the contention that the feedback had temporarily altered the internal model.

Others have also used the pitch-shift perturbation technique for the study of the role that auditory feedback plays in control of voice F_0. When voice pitch feedback is unexpectedly either increased or decreased during vowel phonation or speech production, subjects compensate for the perturbation with voice F_0 responses that change in a direction opposite to that of the perturbation as if to nullify an unplanned deviation in F_0 (Elman, 1981; Kawahara, 1995; Burnett *et al.*, 1998; Hain *et al.*, 2000; Natke and Kalveram, 2001; Jones and Munhall, 2002). These compensatory responses have been interpreted as support for the idea that auditory feedback can be used online to help stabilize voice F_0. Other studies have shown that voice F_0 responses to perturbations in voice pitch feedback may also be observed during speech (Donath *et al.*, 2002; Jones and Munhall, 2002; Xu *et al.*, 2004; Chen *et al.*, 2007). Moreover, such responses occur rapidly enough to correct for errors within the same syllable (Xu *et al.*, 2004) or subsequent syllables (Donath *et al.*, 2002; Xu *et al.*, 2004).

Recently, several studies of pitch-shifted voice feedback demonstrated that voice F_0 responses to pitch perturbations could be modulated according to the demands of the vocal task during either vowel phonation (Liu and Larson, 2007) or speech production (Natke *et al.*, 2003; Xu *et al.*, 2004). Task-dependent modulation of such responses was tested more directly in a study where English speakers either spoke a phrase or produced vowel phonations (Chen *et al.*, 2007). During speech production, vocal responses to downward pitch-shift stimuli presented before a rising F_0 contour were larger than to upward stimuli. Furthermore, overall responses to pitch perturbations during speech production were larger than those during vowel phonations. The larger responses to the downward stimuli during speech may indicate that the system recognized that the downward stimuli were in the wrong direction to the planned upward F_0 contour, and a corrective response was needed. Implicit in this assumption is the idea that the F_0 level existing prior to a planned change in F_0 is known or can be predicted from the present F_0 level. If the actual F_0 level at the time of implementation of the change deviates from the predicted level, a response would be generated to correct for the error and to complete the planned change accurately. Thus the system controlling voice F_0 evidently takes into consideration the present F_0 level while planning for future changes in F_0. If the system plans ahead for future changes in F_0, there must be a time period during which this planning takes place.

The present study was designed to determine the boundaries of the temporal window during which planning for a subsequent change in F_0 takes place. Results from the study of Chen *et al.* (2007) suggest that the system is more sensitive to pitch-shift stimuli before a planned change in F_0 takes place. Therefore, measurement of response magnitudes to pitch-shift stimuli presented at different times before a planned change in F_0 may reveal the temporal boundaries of these periods of increased sensitivity to pitch-shift stimuli. In the present study, Mandarin subjects spoke a phrase that contained different intonation patterns while hearing perturbations in their voice pitch feedback. Pitch-shift stimuli were presented at three different times prior to a planned change in F_0 to test the hypothesis that response magnitude or latency would be modulated according to the timing of stimulus presentation in relation to the planned F_0 trajectory.

II. METHODS

A. Subjects

Ten native speakers of Mandarin (3 males and 7 females) between the ages of 24 and 30 were recruited. All subjects reported normal hearing, and none reported a history of speech or language problems or neurological disorder. All subjects signed informed consent approved by the Northwestern University Institutional Review Board and were paid for their participation.

B. Apparatus

Subjects were seated in a sound-treated room. They wore Sennheiser headphones with attached microphone (model HMD 280). The voice signal from the microphone was amplified with a Mackie mixer (model 1202), shifted in pitch with an Eventide Eclipse Harmonizer (H3000, SE), mixed with a 40 dB SPL pink masking noise (Goldline model PN2; spectral frequencies of 1–5000 Hz) through another Mackie mixer (model 1202-VLZ), and further amplified to 10 dB greater at the headphones than at the microphone with a Crown D75 amplifier and HP 350 dB attenuators. This 10 dB gain in feedback over the voice was used to reduce the chances that the subjects heard their nonshifted bone-conducted signal during vocalization. MIDI software (MAX/MSP, Version 4.5 by Cycling'74) on a laboratory computer was used to control the harmonizer and to generate control (TTL) pulses. A Brüel & Kjær sound level meter (model 2250) and in-ear microphones (model 4100) were used to make acoustic calibrations prior to testing any of the subjects. Subjects monitored their voice loudness from a Dorrough loudness monitor (model 40-A), providing a vi-

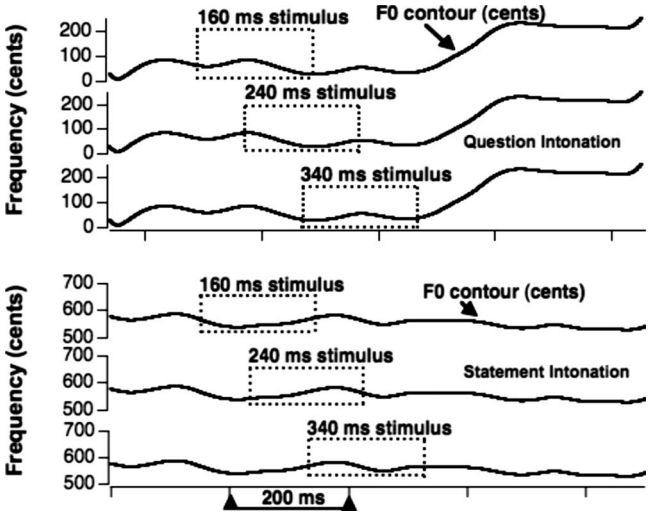


FIG. 1. Illustration of the timing of the stimuli relative to the F_0 contours for the question intonation (top) and statement intonation (bottom). The dashed line boxes indicate the timing and duration of the pitch-shifted stimuli.

sual feedback reference of their voice amplitude. Digitization of the voice output signal, feedback, and control pulses were performed at 10 kHz (12 bits), low-pass filtered at 5 kHz through POWERLAB (model ML880, AD Instruments), and recorded on a computer using CHART software (AD Instruments).

C. Procedures

Subjects were instructed, upon hearing one of two prerecorded speech samples (male voice, ≈80 dB SPL) over the headphones, to repeat them with the same exact inflection pattern within 1 s. The Mandarin sentences consisted of four syllables, "mao1 mi1 mo1 ma1" (kitty touches mom), in which the numeral 1 represents the high-level tone. The pitch on the final syllable "…ma1" was raised in the question intonation and held constant in the statement intonation. The MIDI program, triggered by the onset of the subject's voice, delivered a signal to the harmonizer that modulated the voice pitch feedback of the subject's voice. The variability in the timing of the MIDI output from the onset of the pulse from the vocal detection circuit was about 25 ms. In each block of 60 trials, the model presented to the subjects on each trial had the same intonation pattern, and the stimulus timing was held constant at one of the three values. On each trial, the voice pitch feedback was either shifted up (increasing), down (decreasing), or not changed (control trials). The sequence of stimuli was randomized so that subjects could not predict which type of stimulus would occur on any given trial. Thus the subjects received 20 increases in pitch feedback, 20 decreases, and 20 control trials for each condition. Across six blocks (three stimulus timings by two intonation patterns) of 60 trials each, the stimulus magnitude was fixed at ±100 cents (200 ms duration), but the timing was varied at 160 ms, 240 ms and 340 after vocal onset in the question and the statement intonations, as shown in Fig. 1. Table I lists the averaged timing values of pitch perturbations for three stimuli in the question and statement intonation. During any block of the trials, the stimulus timing was held constant at one of the three delay times. Because of the natural variations in speaking rate, we attempted to keep speaking rate constant by having the subjects repeat a model, even though it was recognized that variation in timing would still occur. We also judged that attempting to measure the syllable durations for each subject would be impractical because here again, rates and syllable durations would vary. We therefore concluded that selecting a single set of delay times was the best compromise between experimental design and variations in production both within and across subjects.

TABLE I. Averaged timing values in milliseconds (SD) of stimuli relative to the onset of syllables 1–4 as a function of intonation, in which positive and negative values, respectively, indicate the presentation of pitch perturbation after and before the utterance of the syllable.

Stimulus (ms)	Intonation	Syllable 1	Syllable 2	Syllable 3	Syllable 4
160	Question	212 (26)	43 (31)	−150 (30)	−369 (38)
	Statement	204 (9)	13 (20)	−207 (28)	−438 (44)
240	Question	292 (25)	128 (22)	−69 (21)	−318 (85)
	Statement	282 (13)	87 (22)	−134 (37)	−381 (43)
340	Question	388 (17)	217 (20)	−39 (14)	−288 (24)
	Statement	387 (19)	196 (16)	−36 (18)	−279 (42)

For data analysis, the voice signal was processed in PRAAT (Boersma, 2001) using an autocorrelation method to produce a train of pulses corresponding to the fundamental period of the voice waveform. This pulse train was then converted into an F_0 analog wave in IGOR PRO (Wavemetrics, Inc., Lake Oswego, OR). An operator manually marked the beginning and end points of the F_0 wave for each sentence according to the voice waveform, and then all the vocalizations in a block were time-normalized (linear interpolation) to reduce temporal variability and enhance the averaging procedures. An average of voice F_0 response was calculated by triggering the averaging computer on the TTL control pulse separately for each condition (phrase type, stimulus onset delay, and stimulus direction), summating all trials and dividing by N to produce an average F_0 response for that particular condition. A "difference wave" was then calculated by subtracting the average control wave from the average test waves derived from the increasing or decreasing pitch-shift stimulus. A point-by-point series of t-tests was run between all control and all test trials for a given condition and subject (Xu *et al.*, 2004). Valid responses were defined by significant t-tests ($p=0.02$) beginning at least 60 ms after the stimulus onset and lasting at least 50 ms. Response latencies were defined as the point where the p values of significant differences decreased below 0.02 and remained decreased for at least 50 ms. Response measurements were made during

TABLE II. Total number of "following" (FOL), "opposing" (OPP), and "non-response" (NR) across stimulus direction.

	Up	Down	Total
OPP	55	50	105
FOL	3	4	7
NR	2	6	8
Total	60	60	120

TABLE III. Total number of following (FOL), opposing (OPP), and non-response (NR) across stimulus timing.

	160 ms	240 ms	340 ms	Total
OPP	33	38	34	105
FOL	1	1	5	7
NR	6	1	1	8
Total	40	40	40	120

the 200 ms period of the stimulus because in the 340 ms condition, the intonated rise in F_0 often began near the end of the 200 ms making it difficult to measure the response. Response magnitudes were measured from the difference wave at the time indicated by the most significant p value within the above-defined measurement window. Statistical analyses were performed on response latency and magnitude using repeated-measures ANOVAs (SPSS, Version 16.0). Assumptions of compound symmetry and circularity for a repeated-measures ANOVA were met. For statistical analysis, non-responses were replaced by the mean value calculated from the measured data from other subjects for that condition.

III. RESULTS

From ten subjects across three stimulus timings, two stimulus directions, and two intonation patterns, there were 120 possible responses ($10\times3\times2\times2$). Tables II–IV list the numbers of opposing (the voice response change was in the opposite direction to the stimulus), "following" (response and stimulus changes in the same direction), and non-responses across stimulus direction, stimulus timing, and intonation pattern, respectively. 88% of the responses "opposed" the stimulus direction. Only 8 out of 120 responses did not meet our criteria of valid response (see Sec. II C) and were declared to be non-responses. The number of valid responses did not vary greatly across the experimental conditions.

Figures 2 and 3 illustrate average responses from representative subjects to pitch-shifted stimuli (thick black lines with error bars) superimposed on average control curves (thin black lines with error bars) across three stimulus timings in the question and statement intonations, respectively. Responses to the upward stimuli are illustrated on the left and downward stimuli on the right. The square brackets at the bottom of these figures indicate the time and direction of the stimulus. The major increase in traces in Fig. 2 coincides with the onset of the fourth syllable. The separation between the control (thin lines) and test (thick lines) wave that begins during the stimulus is the response to the perturbation. Compensatory vocal responses were produced across all experimental conditions in Figs. 2 and 3, except for the upward response for the 340 ms timing in the question intonation (bottom left in Fig. 2) that is a following response. In Fig. 2 (question intonation), it appears that response magnitudes for the responses in the 340 ms timing condition (right) are smaller than those for the 160 and 240 timing conditions. All the responses started before the rising of the F_0 contour of the last syllable. In Fig. 3 (statement intonation), there are no obvious differences in response magnitude across the three stimulus timing conditions.

Figures 4 and 5 show box plots of response magnitudes and latencies as a function of stimulus timing and direction in the question and statement intonations, respectively. Values of response magnitude and latency across stimulus tim-

TABLE IV. Total number of following (FOL), opposing (OPP), and non-response (NR) across intonation pattern.

	Question	Statement	Total
OPP	48	57	105
FOL	5	2	7
NR	7	1	8
Total	60	60	120

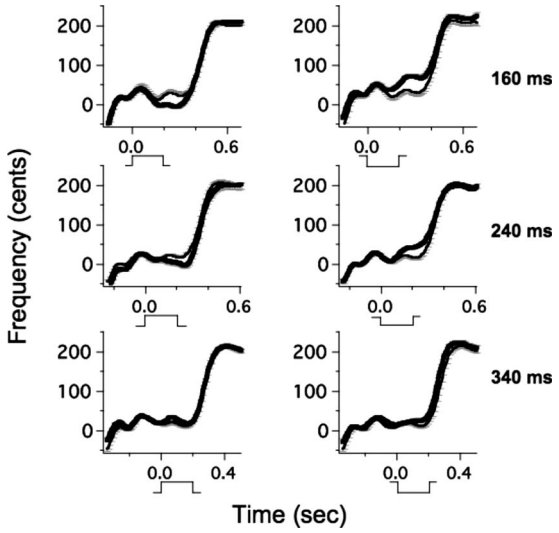


FIG. 2. Average F_0 contours for the test (thick lines) and control (thin lines) F_0 contours for stimulus timings of 160, 240, and 340 ms in the question pattern. Error bars represent the standard error of the mean for a single direction. Square brackets at the bottom indicate the time and the direction of the stimulus.

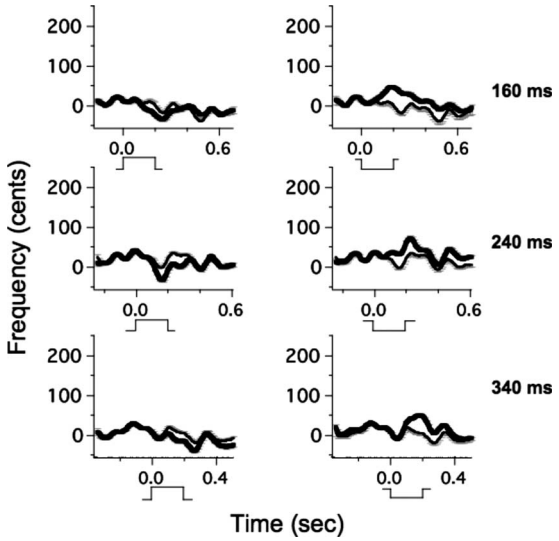


FIG. 3. Average F_0 contours for the test (thick lines) and control (thin lines) F_0 contours for stimulus timings of 160, 240, and 340 ms in the statement pattern. Square brackets at the bottom indicate the time and the direction of the stimulus.

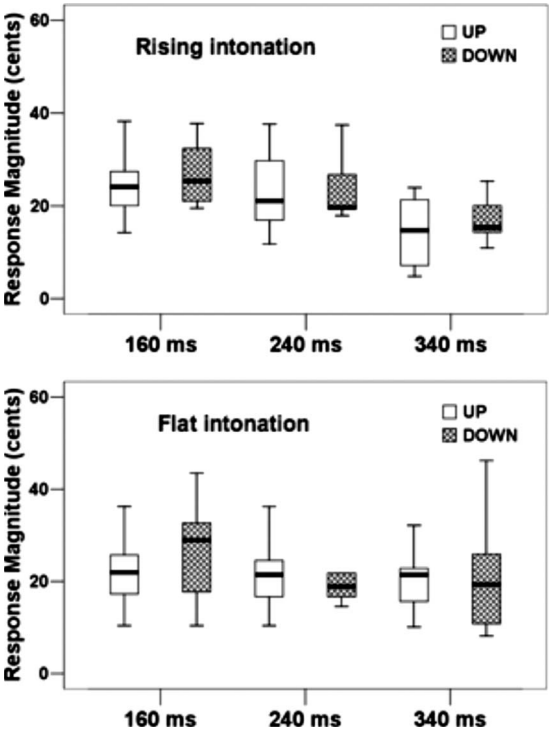


FIG. 4. Box plots illustrating the response magnitude as a function of stimulus timing and stimulus direction in the rising and the flat intonations. Box plot definitions: middle line is median, top and bottom of boxes are 75th and 25th percentiles, and whiskers extend to limits of main body of data defined as high hinge +1.5 (high hinge—low hinge) and low hinge −1.5 (high hinge—low hinge).

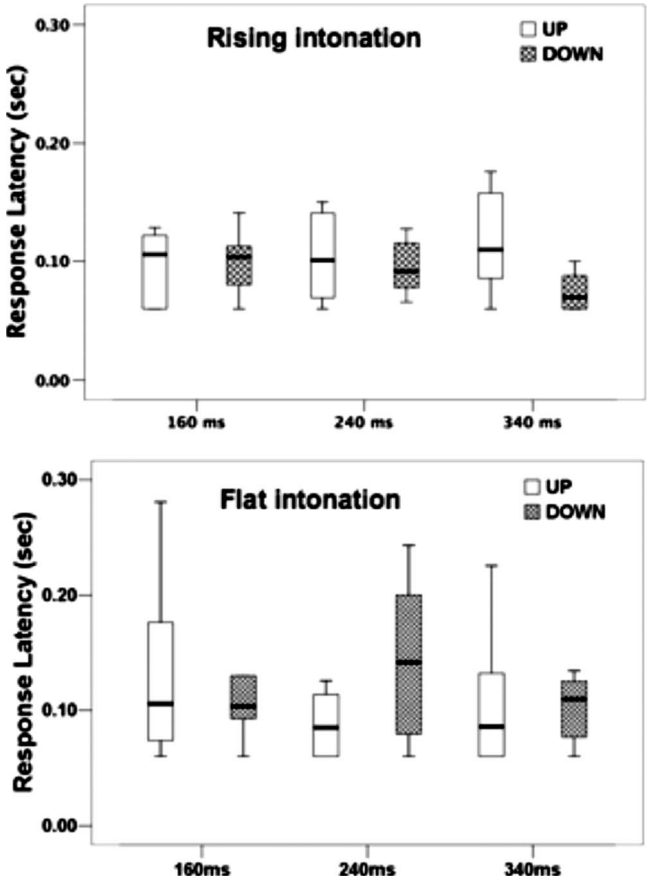


FIG. 5. Box plots illustrating the response latency as a function of stimulus timing and stimulus direction in the rising and the flat intonations.

ing and intonation pattern are shown in Tables V and VI. Three-way repeated-measures ANOVAs (stimulus direction, stimulus timing, and intonation pattern) performed on the response magnitude revealed significant main effects for stimulus timing $[F(2,18)=6.979, p=0.006]$ but not for stimulus direction $[F(1,9)=2.083, p=0.183]$ or intonation pattern $[F(1,9)=0.008, p=0.993]$. A significant interaction effect was found between intonation pattern and stimulus timing $[F(2,18)=6.826, p=0.006]$. In order to determine the effect of stimulus timing on vocal responses in the different intonations, two-way repeated-measures ANOVAs were performed on the response magnitude and latencies for the rising and the flat intonation patterns. The results indicated significant main effects across stimulus timing in the rising $[F(2,18)=9.913, p=0.001]$ but not in the flat pattern $[F(2,18)=1.276, p=0.303]$. No significant main effects were found for stimulus direction in the question $[F(1,9)=0.286, p=0.625]$ or the statement condition $[F(1,9)=7.028, p=0.250]$, and there were no interaction effects between stimulus direction and stimulus timing for the intonation patterns [question: $F(2,18)=0.390, p=0.768$; statement: $F(2,18)=3.236, p=0.180$]. *Post hoc Bonferroni* tests in the rising intonation indicated that the 160 ms (26 ± 7 cents) and 240 ms (23 ± 9 cents) conditions produced significantly larger response magnitudes than the 340 ms (16 ± 6 cents) condition ($p=0.001$ and $p=0.014$, respectively). In addition, three-way repeated-measures ANOVAs performed on the response latency revealed no significant main effects across all conditions.

IV. DISCUSSION

The purpose of the present study was to investigate the effect of timing of pitch-shifted voice feedback on the vocal response across two different intonation patterns during meaningful Mandarin speech. The findings indicated significantly larger response magnitudes for the 160 and 240 ms stimulus onset conditions than for the 340 ms condition in the question intonation, while no significant differences existed between 160 and 240 ms timing conditions. Furthermore, there was no effect of stimulus timing on the responses in the statement intonation.

It is now well known that perturbations in voice pitch feedback evoke compensatory changes in voice F_0 during sustained vowels, singing and speaking conditions (Elman, 1981; Kawahara, 1995; Burnett *et al.*, 1998; Hain *et al.*, 2000; Natke and Kalveram, 2001; Donath *et al.*, 2002; Jones and Munhall, 2002; Xu *et al.*, 2004; Chen *et al.*, 2007). The mechanisms of these responses are unknown but necessarily

TABLE V. Average response magnitudes in cents (SD) as a function stimulus timing and intonation pattern, in which values were collapsed across the stimulus direction.

	160 ms	240 ms	340 ms	Mean
Rising	26 (7)	23 (9)	16 (6)	21 (9)
Flat	23 (9)	22 (8)	21 (9)	22 (9)
Mean	24 (8)	23 (9)	19 (8)	22 (9)

TABLE VI. Average response latencies in milliseconds (SD) as a function stimulus timing and intonation pattern, in which values were collapsed across the stimulus direction.

	160 ms	240 ms	340 ms	Mean
Question	112 (43)	98 (30)	100 (45)	104 (39)
Statement	125 (60)	114 (56)	102 (42)	113 (53)
Mean	118 (52)	106 (45)	101 (43)	109 (47)

involve detection of a change in voice pitch feedback, comparison of the feedback with an internal referent of the desired F_0, activation of neurons related to motor output, transmission delays in the central and peripheral nervous system, and contraction (or relaxation) of laryngeal muscles leading to a change in voice F_0. Question intonation in Mandarin involves a large increase in F_0 at the utterance-final position (Ho, 1977; Lin, 2004; Liu and Xu, 2005), although smaller F_0 increases also occur before the final syllable (Yuan *et al.*, 2002; Liu and Xu, 2005; Ni *et al.*, 2006). For statements, it has been found that there is an F_0 decline over the course of an utterance in Mandarin (Xu, 1999; Shih, 2000), but the declination is very slight if the utterance consists of only the high-level tone (Xu, 1999). These patterns are also found in the utterances produced by our subjects. Large F_0 differences occurred at the utterance-final position between questions (288 Hz) and statements (241 Hz), $[F(1,9)=51.593, p=0.000]$ for this mix of male and female subjects. Smaller F_0 differences also occurred in the first syllable, 257 Hz in questions and 240 Hz in statements $[F(1,9)=5.753, p=0.040]$.

When a speaker plans on increasing voice F_0 at the end of an utterance in a question, as is the case in the present study, there must be a planning process that is aware of the current F_0 level and the correct time to contract muscles to change F_0. The observations of the present study suggest that there is a time-dependent interaction between responses to pitch perturbations and mechanisms involved in planning a question intonation in Mandarin speech. The interaction is time-dependent because only responses to the 340 ms stimuli, not responses to stimuli that occurred earlier, were affected. The interaction is related to the planning of the increase in F_0 because the reduction in response magnitude only occurred with the question intonation condition, not the statement intonation condition.

Further observations suggest the dimensions of the time window during which planning for the F_0 contour may take place. With the 340 ms stimulus onset time, the response began about 100 ms following the stimulus onset and about 100 ms before the rising F_0 contour of the fourth syllable. The decrement in response magnitude was not observed for the 160 or 240 ms conditions with the rising intonation (question). Therefore, the reduction in the response is related to the rising intonation on the last syllable. Moreover, the time period when the interaction between the pitch-shift stimulus and the decrement in response magnitude occurred must be between the stimulus onset, 200 ms before the rising intonation, and the beginning of the response, 100 ms before the rising intonation. It is therefore suggested that there is a time window of 100 ms duration (between 200 and 100 ms before a planned increase in F_0 contour) during which auditory feedback may interact with the central planning mechanisms.

It is next instructive to consider the reasons for this interaction. One explanation for the reduction in response magnitudes observed for the 340 ms condition with the rising F_0 contour may relate to the pattern of voice F_0 regulation in question intonation. When a sentence is spoken as a question with either final or neutral focus, F_0 starts to rise exponentially from the start of the last word (Liu and Xu, 2005). In the present study, the final word is the syllable /ma/ for "mother." In other words, the F_0 of /ma/ is controlled by a lexical function as well as an intonation function. Such dual control presumably requires greater amounts of neural resources than when only a single function is involved. This control may have depleted some of the neural resources needed to monitor sensory, including auditory feedback.

Another possible explanation for these results is that the auditory perturbation interacted with mechanisms that were involved in elevating F_0 at the end of the sentence. One type of interaction may be that the mechanisms responsible for the voluntary increase in F_0 caused a reduction in sensitivity to the perturbation so that the response to the perturbation would not interfere with the planned increase in F_0. It is also possible that there was active attenuation of the responding mechanisms, again possibly to avoid interference with the planned rise in F_0. Regardless of the exact mechanisms and the reasons for the response reduction, the results suggest that there is a time window (100–200 ms) before a rise in F_0 during which a response to an auditory perturbation may interact with speech planning mechanisms.

Another interpretation of the data relates to the method of measurement of the responses. Due to the fact that for the question intonation condition, there was a rising intonation on the fourth syllable and the need to make response measurement conditions consistent across all conditions, it was necessary to limit the time window for measurement of responses in all conditions to the 200 ms stimulus window. If it had been possible to extend the measurement window out further in time, response magnitudes may have been the same in all timing conditions. Nevertheless, the fact that the responses in the 340 ms condition were smaller than those for the 160 and 240 ms conditions within a 200 ms time period following the stimulus indicates that response-generating mechanisms behaved differently in the 340 ms condition compared with the other timing conditions.

A. Comparison with other studies

The effect of stimulus timing on the response during the production of bi-tonal Mandarin phrase was also investigated by Xu *et al.* (2004). The disyllabic nonsense phrases /ma ma/ with different tonal patterns (high-high, high-rising, and high-falling) were tested with stimuli of 100 and 250 ms delays after the onset of the voice. The results indicated no effect of stimulus timing on the response magnitude or the latency. This finding is in contrast to the present study and may be explained by the fact that only tonal changes were manipulated in the study of Xu *et al.*, without intonation

changes, which may require a different response mechanism during the dynamic control of voice F_0. Another explanation is that a nonsense phrase was used in the study of Xu *et al.* (meaningful words in a nonsense phrase), and meaningful speech was tested in the present study. In the present study, however, data in the statement intonation indicated no effect of stimulus timing on the response, which is comparable with the findings of Xu *et al.* in the high-high (/ma ma/) condition.

In comparing this study with previous pitch shift experiments, the variability in speech style (meaningful speech vs nonsense speech) may contribute to the variability in response magnitudes and latencies. For example, during nonsense speech, vocal response magnitudes of 47 cents (Natke *et al.*, 2003) were reported for 100 cents stimuli, and magnitudes of 50 cents (static tone) and 85 cents (dynamic tone) were reported for 200 cents stimuli (Xu *et al.*, 2004). These values are larger than the overall response magnitudes during meaningful speech of 22 cents in the present study (see Table V). The smaller magnitude responses reported for meaningful speech compared to nonsense speech in these studies may be partially explained by findings from the present study that intonation pattern and stimulus timing can modulate response magnitudes. It should come as no surprise, therefore, if responses to pitch-shifted feedback during some specific types of speech production conditions may be smaller than those in nonsense speech.

One problem with the interpretation of the data in the present study comes from a comparison of the present results with those reported by Chen *et al.* (2007). In that study, responses to 200 cents downward stimuli that were presented prior to a rise in F_0 during English speech were larger than those to upward stimuli. In the present study, there was no difference in magnitude of responses between upward and downward stimuli in any of the timing conditions. The 340 ms condition was most similar to that tested in the study of Chen *et al.*, and it is surprising that in the present study responses to the downward stimuli were not larger than those to the upward stimuli, as they were in the study of Chen *et al.* Two factors may explain this. First, although the rise in F_0 at the end of the phrase was a supra-segmental adjustment in both studies, the Mandarin language background of the subjects in the present study may have predisposed them to respond differently than the English speakers in the study of Chen *et al.* A recent finding has shown that speakers of English produce a much larger final rise in question intonation than do speakers of Mandarin (Liu and Xu, 2007). Second, the effect of the increase in response magnitude in the study of Chen *et al.* was only seen with 200 cents stimuli, not 100 cents stimuli as used in the present study. That is, the stimulus magnitude in the present study may have been too small to evoke the directional difference in the auditory system, causing no significant differences between the responses as a function of stimulus direction.

With respect to response latency, mean values of 211 ms (Jones and Munhall, 2002), 165 ms (Xu *et al.*, 2004), and 150 ms (Natke *et al.*, 2003) were reported for nonsense speech, while 122 ms (Chen *et al.*, 2007) and 109 ms (see Table VI) were observed in the present study during meaningful speech. One explanation for these differences may be that they are attributable to methodologies. Jones and Munhall used a value of 2 standard deviations (SDs) above or below the pre-stimulus mean to define the onset of a response. Natke *et al.* used the Castellan change-point test to define the "change point" in F_0 contours following a stimulus. The methods for measuring response latency in the present study are identical to those used in the studies of Xu *et al.* (2004) and Chen *et al.* (2007) and are hence more comparable to those studies.

Other than the methodological differences, the latency differences suggest that responses to pitch-shifted voice feedback may be faster during meaningful speech than during nonsense speech. This observation may be explained by considering that the target of natural speech is to communicate with listeners. Even though response magnitudes during meaningful speech may be smaller than those during nonsense speech in some conditions, it may be more important for communication effectiveness for subjects to correct for errors faster during meaningful speech than during nonsense speech, resulting in the shorter latencies during meaningful speech than during nonsense phrases.

V. CONCLUSION

In the present study, the effect of the stimulus timing on the pitch-shift reflex during Mandarin speech was investigated in different intonation patterns. The results revealed that stimulus timing is an important variable that can affect the magnitude of vocal responses to pitch perturbations when they occur in a question intonation. Responses to pitch perturbations that were presented just prior to a planned pitch increase were reduced in magnitude compared to stimuli that were presented earlier or to those that were presented when there was no planned F_0 change. The reduction in response magnitude to stimuli that are presented just before a planned increase in voice F_0 demonstrates a contrast to previous reports that showed an increase in response magnitude in some speaking conditions. In combination, these contrasting effects demonstrate that the audio-vocal system can regulate the vocal response depending on the specific task. The factors that appear to be responsible for these differing effects are the timing of the stimulus and whether or not a subsequent change in F_0 is planned. Future studies directed at defining the time window in which stimuli can lead to changes in response magnitude may help us to understand the timing of neural mechanisms involved in the control of temporal features of speech.

ACKNOWLEDGMENT

This work was supported by NIH Grant No. 1R01DC006243. We thank Mr. Chun Liang Chan for programming assistance and Ms. Dilpreet Kaur Minhas for her assistance in data analysis.

Angelocci, A. A., Kopp, G. A., and Holbrook, A. (**1964**). "The vowel formants of deaf and normal-hearing eleven- to fourteen-year-old boys," J. Speech Hear. Disord. **29**, 156–170.

Binnie, C. A., Daniloff, R. G., and Buckingham, H. W. (**1982**). "Phonetic disintegration in a five-year-old following sudden hearing loss," J. Speech

Hear. Disord. **47**, 181–189.
Boersma, P. (**2001**). PRAAT, a system for doing phonetics by computer (Glot International).
Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice *F*0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.
Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (**2007**). "Voice *F*0 responses to pitch-shifted voice feedback during English speech," J. Acoust. Soc. Am. **121**, 1157–1163.
Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). "Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables," J. Acoust. Soc. Am. **111**, 357–366.
Eady, S. J., and Cooper, W. E. (**1986**). "Speech intonation and focus location in matched statements and questions," J. Acoust. Soc. Am. **80**, 402–416.
Elliott, L., and Niemoeller, A. (**1970**). "The role of hearing in controlling voice fundamental frequency," Int. Audiol. **9**, 47–52.
Elman, J. L. (**1981**). "Effects of frequency-shifted feedback on the pitch of vocal productions," J. Acoust. Soc. Am. **70**, 45–50.
Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–141.
Ho, A. T. (**1977**). "Intonation variation in a Mandarin sentence for three expressions: Interrogative, exclamatory and declarative," Phonetica **34**, 446–457.
Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of *F*0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.
Jones, J. A., and Munhall, K. G. (**2002**). "The role of auditory feedback during phonation: Studies of Mandarin tone production," J. Phonetics **30**, 303–320.
Kawahara, H. (**1995**). "Hearing voice: Transformed auditory feedback effects on voice pitch control," in 'Computational Auditory Scene Analysis' and 'International Joint Conference on Artificial' Intelligence, Montreal.
Kawato, M. (**1999**). "Internal models for motor control and trajectory planning," Curr. Opin. Neurobiol. **9**, 718–727.
Lane, H., Wozniak, J., Matthies, M., Svirsky, M., Perkell, J., O'Connell, M., and Manzella, J. (**1997**). "Changes in sound pressure and fundamental frequency contours following changes in hearing status," J. Acoust. Soc. Am. **101**, 2244–2252.
Larson, C. R., Altman, K. W., Liu, H., and Hain, T. C. (**2008**). "Interactions between auditory and somatosensory feedback for voice F_0 control," Exp. Brain Res. **187**, 613–621.
Leder, S. B., Spitzer, J. B., and Kirchner, J. C. (**1987**). "Speaking fundamental frequency of postlingually profoundly deaf adult men," Ann. Otol. Rhinol. Laryngol. **96**, 322–324.
Lin, M. (**2004**). "On production and perception of boundary tone in Chinese intonation," in *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages* (Chinese Academy of Social Sciences, Beijing), pp. 125–129.
Liu, F., and Xu, Y. (**2005**). "Parallel encoding of focus and interrogative meaning in Mandarin intonation," Phonetica **62**, 70–87.
Liu, F., and Xu, Y. (**2007**). "Question intonation as affected by word stress and focus in English," in *Proceedings of the 16th International Congress of Phonetic Sciences*, edited by J. Trouvain and W. J. Barry (International Congress of Phonetic Sciences, Saarbrücken), pp. 1189–1192.
Liu, H., and Larson, C. R. (**2007**). "Effects of perturbation magnitude and voice F_0 level on the pitch-shift reflex," J. Acoust. Soc. Am. **122**, 3671–3677.
Mallard, A. R., Ringel, R. L., and Horii, Y. (**1978**). "Sensory contributions to control of fundamental frequency of phonation," Folia Phoniatr. Logop. **30**, 199–213.
Metzner, W. (**1996**). "Anatomical basis for audio-vocal integration in echolocating horseshoe bats," J. Comp. Neurol. **368**, 252–269.
Mürbe, D., Pabst, F., Hofmann, G., and Sundberg, J. (**2002**). "Significance of auditory and kinesthetic feedback to singers' pitch control," J. Voice **16**, 44–51.
Natke, U., Donath, T. M., and Kalveram, K. T. (**2003**). "Control of voice fundamental frequency in speaking versus singing," J. Acoust. Soc. Am. **113**, 1587–1593.
Natke, U., and Kalveram, K. T. (**2001**). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," J. Speech Lang. Hear. Res. **44**, 577–584.
Ni, J., Kawai, H., and Hirose, K. (**2006**). "Constrained tone transformation technique for separation and combination of Mandarin tone and intonation," J. Acoust. Soc. Am. **119**, 1764–1782.
Oller, D., and Eilers, R. (**1988**). "The role of audition in infant babbling," Child Dev. **59**, 441–449.
Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., and Guiod, P. (**1997**). "Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models," Speech Commun. **22**, 227–249.
Pierrehumbert, J., and Hirschberg, J. (**1990**). in *Intentions in Communication*, edited by P. R. Cohen, J. Morgan, and M. E. Pollack (Massachusetts Institute of Technology Press, Cambridge, MA), pp. 271–311.
Shih, C. (**2000**). in *Intonation: Analysis, Modelling and Technology*, edited by A. Botinis (Kluwer Academic, Heidelberg), pp. 243–268.
Sinnott, J. M., Stebbins, W. C., and Moody, D. B. (**1975**). "Regulation of voice amplitude by the monkey," J. Acoust. Soc. Am. **58**, 412–414.
Smotherman, M. S. (**2007**). "Sensory feedback control of mammalian vocalizations," Behav. Brain Res. **182**, 315–326.
Sundberg, J., Iwarsson, J., and Billstrom, A. H. (**1995**). "Significance of mechanoreceptors in the subglottal mucosa for subglottal pressure control in singers," J. Voice **9**, 20–26.
Svirsky, M. A., Lane, H., Perkell, J. S., and Wozniak, J. (**1992**). "Effects of short-term auditory deprivation on speech production in adult cochlear implant users," J. Acoust. Soc. Am. **92**, 1284–1300.
Waldstein, R. (**1990**). "Effects of postlingual deafness on speech production: Implications for the role of auditory feedback," J. Acoust. Soc. Am. **88**, 2099–2114.
Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (**1995**). "An internal model for sensorimotor integration," Science **269**, 1880–1882.
Wolpert, D. M., Miall, R. C., and Kawato, M. (**1998**). "Internal models in the cerebellum," Trends Cogn. Sci. **2**, 338–347.
Xu, Y. (**1999**). "Effects of tone and focus on the formation and alignment of F_0contours," J. Phonetics **27**, 55–105.
Xu, Y., Larson, C., Bauer, J., and Hain, T. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.
Xu, Y., and Xu, C. X. (**2005**). "Phonetic realization of focus in English declarative intonation," J. Phonetics **33**, 159–197.
Yuan, J., Shih, C., and Kochanski, G. P. (**2002**). "Comparison of declarative and interrogative intonation in Chinese," in *Proceedings of the First International Conference on Speech Prosody* (International Conference on Speech Prosody, Aix-en-Provence, France), pp. 711–714.

Place assimilation across words in running speech: Corpus analysis and perception

Frank Zimmerer
Goethe Universität, Institut für Phonetik, Georg-Voigt–Str. 6/2, 60325 Frankfurt, Germany and Universität Konstanz, FB Sprachwissenschaft, D186, 78457 Konstanz, Germany

Henning Reetz
Goethe Universität, Institut für Phonetik, Georg-Voigt–Str. 6/2, 60325 Frankfurt, Germany

Aditi Lahiri[a)]
Universität Konstanz and Centre for Linguistics and Philology, University of Oxford, Walton Street, Oxford OX1 2HG, United Kingdom

(Received 18 July 2007; revised 2 August 2008; accepted 20 October 2008)

Place assimilation can lead to neutralization of segmental contrasts. It is controversial, however, to what extent such neutralizations actually happen in natural speech. This study examines: (i) the degree to which regressive place assimilations occur in word final consonants in conversational German, and (ii) whether these assimilations are perceived as neutralized by listeners. The production analysis, based on spontaneous speech, shows that complete assimilations do take place in conversational speech and that there is a clear asymmetry between coronal versus labial and dorsal segments. Furthermore, function words show a higher degree of assimilation than lexical words. Two experiments examined the effects of assimilation on perception. A forced choice reaction time perception experiment, using nasal stimuli from the corpus, examined how fast and accurately listeners identified sounds in different segmental contexts. Results indicate that (a) with equal accuracy and speed, listeners identified original and assimilated [m]s; (b) unassimilated-/m/s were identified equally well across contexts, but not unassimilated-/n/s. A free transcription experiment reproduced these findings. An acoustic analysis provides further evidence that regressive place assimilation across word boundaries can result in absolute neutralization of place contrasts in running speech. The results support models predicting asymmetries between coronal versus labial and dorsal consonants. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3021438]

PACS number(s): 43.70.Mn, 43.71.Es, 43.71.Sy [PEI] Pages: 2307–2322

I. INTRODUCTION

Speech is variable—certainly across speakers and across dialects, but also within a single speaker depending on speech rate or style. Acoustic shapes of lexical items easily vary across contexts or sentence types. Within a model of discrete phonological entities, it is customary to assume that reductions, deletions, insertions, and assimilations are processes which modify or alter idealized lexical representation in running speech. A German word like *einverstanden* "agree-past participle" occurring 47 times in the Kiel corpus (IPDS, 1994) has 23 different variants in the database (for a complete list see Appendix A).[1]

There is no utterance in the database which exactly matches the canonical pronunciation (i.e., [ˈʔaɪnfɐˌʃtandən]), most cases exhibiting more than one deviation. Not only are there many types of variation, but the deviations from the norm are optional and need not be complete and may still be perceptible. Remnants of a deleted sound may still be present as in [ˈaɪnfɐˌʃtan̰n], [ˈaɪnfɐˌʃtanʔn], or [ˈaɪnfɐˌʃtan̰], where some segments have been deleted completely, or glottalization indicates that a stop (i.e., [d]) has been severely reduced.[2] Transcriptions of place assimilations suggest complete neutralization of a featural contrast, as in [ˈaɪmfɐˌʃtan̰n] where additionally to deletions, the [n] is assimilated to the labiality of [f].[3]

However, as in deletions, traces of the original [n] may still be found in the signal. For instance, Nolan (1992) argued that assimilations were more likely to be gradient than complete, and that target information was available in assimilated sequences (see also Gow, 2002; for voicing assimilation, see Snoeren *et al.*, 2006). Listeners are sensitive to these gradient assimilations in production and could identify residual alveolar gestures in 40% of the assimilated tokens (Nolan, 1992, p. 271). Gow's results also indicate that listeners use the information of the underlying place of articulation even in segments that auditorily sound as if they are completely assimilated (Gow, 2002; Dilley and Pitt, 2007; also Manuel, 1995; Snoeren *et al.*, 2006). Indeed, some researchers express doubt concerning the very existence of complete assimilation (Gow, 2002; see also Snoeren *et al.*, 2006).

However, in a recent extensive coverage of regressive assimilation of naturally spoken American English (Buckeye Corpus of Conversational Speech, Pitt *et al.*, 2006), Dilley and Pitt (2007) found that 9% of coronal (alveolar) word final stops and nasals were transcribed as assimilated to the place of articulation of the following consonant (labials and

a)Electronic mail: aditi.lahiri@ling-phil.ox.ac.uk

velars).[4] Acoustic measurements consisting of the change in F2 and amplitude of the preceding vowel showed that these frequently did not differ between the assimilated consonants and the canonical labials and velars. They conclude that "assimilation is often complete or nearly complete in spontaneous speech" (Dilley and Pitt, 2007, p. 2350). One must note, however, that the F2 values were gradient for both consonants labeled as assimilated, as well as those in an assimilatory context (i.e., followed by labials or velars) as compared to alveolars in a nonassimilatory context (i.e., followed by other alveolars). As the authors report, a possibility exists that the real number of assimilations is underestimated, since even some instances of those that were labeled as unassimilated could be actually assimilated because the labelers are always reasonably conservative.

Thus, in spite of the optional and gradient nature of fast speech processes, neutralization due to assimilation can be perceived as complete. Furthermore, such assimilations can even lead to orthographic changes. Orthography tends to be conservative and even if a pronunciation change has occurred, the spelling remains often unaltered. However, when the orthography changes (without formal institutional intervention such as the German *Rechtschreibreform*), one is reasonably sure that a change has really taken place. For instance, words with the negative prefix {in-} have been borrowed into English from Romance at different times. A word like *impossible* could be spelt earlier as ⟨inpossible⟩: *It es bot foli al pi talking, And als an inpossibile thing* **1300** *Cursor M.* 14761 (OED, 1989, p. 732). The ⟨n⟩ is now always pronounced as a labial and this place assimilation changing the [n] of {in-} to [m] when labial [p, b, m] follow is now also always reflected in spelling.[5] Listeners must have perceived the assimilation to lead to a change in the (conservative) orthography. A new formation like *input*, which is not made with the negative prefix, preserves the ⟨n⟩ in the spelling, although it is also pronounced with a [m].

Although there is little doubt that dictionary-like pronunciations are not the norm in connected speech and there is an increasing number of spoken language corpora which are used in publications, such as Snoeren and colleagues (2006) and Dilley and Pitt (2007), there is still a dearth of statistically reliable data as to what extent connected speech phenomena like assimilations actually occur in other languages and moreover, even less is known about how they are perceived by normal listeners and trained labelers. In this paper, we analyze normal running speech to quantify how often the contexts for assimilations arise, how frequently such assimilations are realized, and how they are perceived, focusing on German where there is no study as yet bringing together corpus analysis and the repercussions for perception. In the following sections we first discuss possible across-word assimilation phenomena in German and provide a complete analysis of such assimilations in the Kiel corpus. Our analysis is first divided into a section for function words and a separate section for lexical words. There is ample evidence that the phonological and phonetic behavior of these two word categories is different (e.g., Selkirk, 1984; Kaisse, 1985; Nespor and Vogel, 1986; Hall, 1999; Ogden, 1999; Philipps, 2001; Local, 2003; Kabak and Schiering, 2006; Bybee, 2007). For instance, function words in German are often drastically reduced (Hall, 1999). For English, it has been reported that /m/ in the function words "I'm" may assimilate to neighboring segments, whereas /m/ in content words such as "time" does not (Ogden, 1999; Local, 2003). We also examine whether function words behave differently when it comes to regressive place assimilation and compare the findings of the two sections. This analysis is followed by two perception experiments examining the ability of listeners to hear the altered sounds based on the Kiel transcriptions. We offer a proposal for modeling alterations by assimilations in a framework assuming underspecification, such as the Featurally Underspecified Lexicon model (FUL) (Lahiri and Reetz, 2002).

II. ASSIMILATIONS IN GERMAN

How inviolable are word final consonants? Such consonants are of course subject to change, either by progressive assimilation within a word or by regressive assimilation across words. Although progressive place assimilations are reported to be frequent within a word in German, cf. *geben* [gebən]>[gebm], regressive assimilation across words is more controversial (Wurzel, 1970; Dressler *et al.*, 1972; Benware, 1986; Hall, 1992; Wiese, 1996; but see Kohler, 1995). As Wiese (1996) states, when it is possible to pronounce two words as a single unit then regressive assimilation is more likely (cf. *man kommt* "one comes" pronounced as [maŋ kɔmt] versus no assimilation in *der Mann kommt* "the man comes" [man kɔmt]). More definite conclusions regarding regressive assimilations in German is difficult since in his words, "…first, there is little systematic study of such differences, and, second, at the tempo of fast speech, assimilation is certainly possible in the latter example" (Wiese, 1996, p. 221). Nonetheless, regressive assimilations across words are not unknown and the possibility is at least mentioned by most of these authors.[6]

Kohler, however, explicitly claims that regressive place assimilation takes place across word boundaries (Kohler, 1995, p. 206; see also Kohler, 1990) and cites several examples where such assimilations occur. One such example is *bunt machen* "to make colorful" [bʊnt maxŋ] being pronounced as [bʊmp maxŋ]. A study on the Viennese variety of German by Dressler and his colleagues also mentions the possibility of regressive place assimilation in fast speech (Dressler *et al.*, 1972).

Given that there is little systematic work on regressive assimilation across word boundaries in conversational German, we turned to the Kiel corpus for natural speech data. The Kiel corpus (IPDS, 1994) provides us with ideal data for examining minute phonetic variations in naturally spoken dialogues. The corpus makes available detailed phonetic transcriptions and hand labeled segmentation of the acoustic signal performed by trained phoneticians. Thus, we have rich material to study across-word variations as noted by phoneticians who used both the information in the signal as well as their own perception to make decisions concerning what was actually produced.

In what follows, we refer to the sequence of consonants across word boundaries as C_1 and C_2. The word final segment (C_1), which could assimilate, will be referred to as the target and the word initial segment of the following word (C_2) as the trigger as shown in (1). C_1 could be any stop, fricative or nasal in German, whereas C_2 could be any obstruent or nasal which may occur in that position. Word finally (target position), voiced stops and fricatives are devoiced regularly in German (*Auslautverhärtung*—final devoicing—see Kohler, 1995; Wiese, 1996; Hall, 2000, and references therein). Consequently, in production, there are no word-final voiced obstruents.[7]

(1) TARGET (C_1) and TRIGGER (C_2) in word sequences
$[\ldots C_1]_{\omega 1}[C_2 \ldots]_{\omega 2}$
| |
e.g., *Termin bleiben* ("appointment stick"
—stick to appointment)

The following issues are addressed in the analyses of the speech data: (a) How often do German speakers produce regressive assimilations across words? (b) Is there a particular place of articulation for C_1 which favors assimilation? For instance, are [coronal] sounds more likely to assimilate than [labial] ones? (c) Does the manner of articulation of C_1 matter for regressive assimilation? For example, do nasals assimilate more often than stops in running speech? (d) Does the place and manner of articulation of the C_2 trigger correlate with regressive assimilation? (e) Does the lexical status of the first word (function words versus lexical words) increase the probability of assimilation since function words are supposed to be less stable and more vulnerable to alterations?

Other than quantifying the number of assimilations in conversational speech, we also address the issue of completeness of these assimilations in perception. Two perception studies were conducted using selected material from the database. The first experiment used a forced choice phoneme identification task on fragments of words from selected dialogues. In a second experiment, subjects were asked to transcribe freely what they heard. Our goal was to observe how listeners would perceive segments labeled as assimilated in the speech corpus and whether there were any remnants of the original segment to affect the speed and accuracy of identification as compared to unchanged segments.

III. ANALYSIS OF SPONTANEOUS SPEECH

The Kiel corpus consists of about 4 h of dialogues of 42 (northern) German speakers (18 female, 24 male) who were engaged in an appointment-making task. They were each given a schedule with existing and sometimes conflicting appointments, and their task was to decide on future meetings. In order to ensure a high degree of natural speech, the speakers were ignorant of the schedule of their partners. The dialogues were recorded with the speakers placed in different sound-treated rooms communicating by headsets. As mentioned earlier, all dialogues were transcribed and labeled by trained phoneticians using visual scaleable spectrograms and oscillogram displays as well as auditory information (Kohler *et al.*, 1995, p. 33). Other than the phonetic transcription of what was actually pronounced, the corpus contains an orthographic text, along with its canonical phonetic transcription. The idealized canonical transcription denotes how utterances should be if they were spoken in accordance with a careful dictionary-like pronunciation. This allows a direct comparison of a canonical transcription with the actual—phonetically transcribed—pronunciation. The nature of the task and the fact that participants had to make very similar appointments restricted the vocabulary. For instance, dates, times, and days of the week occur very often. Nevertheless, since the speakers were unaware of the purpose of the recordings, the conversations were very natural.

A. Material and methods

Overall the *Kiel Corpus of Spontaneous Speech* consists of 1984 turns of dialogues by 42 speakers available in three CDs. All transcribed dialogs from the appointment-making task consisting minimally of two words, including the test dialogs, went into our analysis. The longest turn is about 54 s and the shortest ones are less than a second. An example of average length of one turn is given in (2) with (a) the orthographic transcription with our English translation, (b) the canonical transcription in IPA, and (c) the phonetic transcription as given in the Kiel corpus (omitting instances of nonspeech sounds such as clicking, smacking, etc.).

(2) Example of a dialogue in the Kiel corpus (utterance g071a006):

(a) Orthographic transcription *Das würde mir ganz ausgezeichnet passen. Machen wir das fest*? "That would be excellent for me. Can we confirm this?,"
(b) Canonical transcription (IPA transcription of the orthographic form) [das vʏɐdə mɪːɐ ˈgants ˈʔaʊsgəˌtsaɪçnət ˈpasən. ˈmaxən viːɐ das ˈfɛst?],
(c) IPA transcription of the actual pronunciation [das vʏɐd mɪːɐ ˈgants ˈãʊsgəˌtsaɪhnət ˈpasn. ˈmaxm viːɐ das ˈfɛst?].

We counted all possible contexts of regressive place assimilations of nasals and obstruents, and then summed up all cases where they actually occurred, even across sentence boundaries. This meant that instances where the C_1-target and C_2-trigger had the same place of articulation were ignored. For the analysis, utterances were excluded where technical problems led to incomplete speech signals, or where speakers produced false starts, or where there were intervening hesitational markers as *ähm* or *m(hm)* ("ahem, hm") like in *machen äh(m) wir*. Furthermore, to rule out possible confounds, we did not include utterances where a possibility of progressive place assimilation existed and thus target and trigger could not be identified unambiguously. For example, the assimilated [m] in a phrase like *haben wir* ("have we") [haːbən viːɐ] spoken as [haːbm viːɐ] has two potential triggers, the preceding labial [b] or the following labial [v], and was therefore not considered in our data set. We also excluded words where the last segment (C_1) was

TABLE I. Obstruents and nasals in German and their phonological place features.

Labial	*bilabial, labiodental*	[m, p, b, f, v, pf]
Coronal	*alveolar, palatoalveolar, palatal*	[n, t, d, s, ʃ, z, ç, ts, tʃ]
Dorsal	*velar*	[ŋ, k, g, x]

deleted to rule out all possible confounds connected to deletions. Thus, phrases like *und Mittwoch* ("and Wednesday") [ʊnt ˈmɪtvɔx] pronounced without word final [t] as [ʊn ˈmɪtvɔx] were not included.[8] All obstruents and nasals were treated as possible triggers (C_2). The phonological features of the consonants that were taken into account, both as target and trigger, are given in Table I.[9]

The segments [ŋ, tʃ, s, x] do not occur word initially in standard German, except in a handful of loanwords. Due to final devoicing, one expects the voiced consonants to be devoiced word finally, but this does not affect place assimilation.

B. Analysis

Function words behave differently than content words in many different ways (e.g., Selkirk, 1984; Kaisse, 1985; Nespor and Vogel, 1986; Hall, 1999; Ogden, 1999; Philipps, 2001; Local, 2003; Kabak and Schiering, 2006; Bybee, 2007). Therefore, we split up the following analysis into a section for function words, another section for content words, and then we compare the results. The issue of interest to us is whether function words behave differently concerning regressive place assimilation. Since there is considerable controversy concerning which words count as function words, we opted for the classification in the Kiel corpus (marked with a final "+" in their transcription).

1. Function words

An overview of the different kinds of function words occurring in the database is given in (3). The function words could be either trigger or target. However, we ignored the syntactic category of the second word in a sequence, since our main point of interest is on the words that undergo assimilation, therefore the lexical status of the trigger was of no relevance in our analysis.

(3) Examples for different function word categories in the Kiel corpus:

(a) Auxiliaries: *bin, hatte, gewesen, möchte* ("am, had, been, would like"),
(b) Determiners: *der, die, das, ein, eine* ("the.*masc*, the.*fem*, the.*neut*, a.*masc*, a.*fem*"),
(c) Pronouns: *ich, wir, Sie, Ihre, Ihnen*, ("I, we, you.*hon*, you.*hon.gen*, you.*hon.acc*")
(d) Prepositions: *in, am, bis*, ("in, at.*dat*, until/to'"),
(e) Demonstratives: *diesen, dieser, diesem* ("this-*case*"),
(f) Conjunctions: *und, aber, zwar* ("and, but, but/namely").

Overall, 4144 function words qualified as target (C_1) in a sequence of two consonants at word boundaries. Out of those, in 266 (6.4%) instances the target C_1 was transcribed in the corpus as having been pronounced with a different place of articulation from the canonical form, e.g., *ein Montag* [aɪm moːntax] instead of [aɪn moːntax]. Table II(a)–(c) show the data for all occurrences of targets and the corresponding triggers, with the numbers and percentages of assimilated segments.[10]

The results clearly indicate that although regressive place assimilations are not obligatory, they do occur across word boundaries of function words. In Table II(b) we see that 232 out of 2961 [coronal] sounds assimilate in place to the following segment, most of which were /n/. Out of a total of 1230 /n/ final function words, 225 or 18.3%, were labeled as assimilated; 168 out of 1036 (16.2%) words ending in /n/ assimilated to [m] and 57 out of 194 (29.4%) changed to [ŋ], when followed by [labial] or [dorsal] consonants, respectively. Out of a total of 200 function words ending in /t/, only 4 assimilated to [p] when a labial followed, and none assimilated before [dorsal] segments. Overall, 1021 function words ended in /s/, one of which was assimilated to a labial [f]. Finally, out of 510 /ç/ final function words, 2 assimilated to [f].

Turning to the [labial] final function words [Table II(a)], there were in all 724 of which 27 assimilated, all of which were /m/. There were 583 instances of /m/ final function words and 27 were labeled as having changed its place of articulation, 23 (i.e., 4.3%) to [n] when followed by a [coronal], 4 (9.3%) to /ŋ/ when followed by a [dorsal]. None of the 82 /p/ or 59 /f/ final function words assimilated. As for the [dorsal] final function words [Table II(c)], they all ended in /x/, and 7 out of 459 instances (1.5%) showed assimilation—six times to [f] when a [labial] followed, and one [s] when a [coronal] consonant followed.

From the data it also becomes evident that there are clear asymmetries in the patterns of assimilation. [Coronal] sounds assimilate more frequently (7.8%) than other places of articulation; cf. [dorsal] (1.5%) and [labial] (3.7%).[11] Another asymmetry concerns the manner of articulation of the targets that undergo assimilation. Nasal sounds are more prone to assimilation than stops, and fricatives assimilate the least.

The question we turn to now is whether these results are special to function words or whether they form a general pattern observable in connected speech.

2. Lexical words

For lexical words, we counted a total of 2916 possible environments for regressive place assimilation. As compared to function words, there were more C_1 [dorsal] segments. Of all possible environments, 127 (4.4%) assimilations were actually realized. An overview over the different targets and triggers is presented in Table III(a)–(c).

The data for lexical words follow a similar assimilation pattern to that of the function words. [Coronal] segments undergo regressive place assimilation in 121 cases, of which 97 were nasals [Table III(b)]. Among the nasals, 8 /n/ (7.4%) were realized as [ŋ]. The rest, i.e., 89 /n/ (9.4%) were produced as [m]. For lexical words, final [t]s accounted for 24 cases (4.5%) of regressive assimilations. Of the 24 instances

where [t] was assimilated, there was one utterance where [t] became [k] (1.8%), 23 cases showed assimilation to [p] (4.9%). No [coronal] fricative changed place of articulation. As for [labial] target segments, there occurred one assimilation: a word final [m] assimilated to [n] preceding a coronal stop [Table III(a)]. No other labial segment assimilated. [Dorsal] segments assimilated 5 times, all of them were [x]; 3 of them assimilated to [labial], 2 to [coronal] [Table III(c)].

Overall, the data of the lexical words also revealed two kinds of asymmetries. First, the nasal consonants assimilated more often than stops or fricatives. The second asymmetry concerns again the place of articulation of the target segment: [coronal] sounds undergo regressive place assimilation much more frequently (6.1%) than [labial] (1.3%) or [dorsal] (0.6%) segments.

C. Comparison of function and lexical words

The general pattern of assimilation was the same for lexical and function words although the latter underwent assimilation more frequently. Overall, we analyzed 7060 possible sequences for regressive place assimilation in the Kiel corpus of which 393 instances of assimilation occurred (see Table IV); i.e., 5.6% of the possible sequences were actually assimilated. Function words assimilated in 266 cases, whereas lexical words account for 127 instances. However, we should also note that function words occurred more often as targets in our sample than lexical words. We counted 4144 function words (58.7%) and 2916 lexical words (41.3%) as targets; hence, 6.4% of the function words and 4.4% of the lexical words assimilated. Nonetheless, function words show a significantly higher degree of regressive assimilation, as a *Chi-Square* test revealed ($\chi^2=13.9$, $p<0.001$).

Figure 1 depicts the relative percentage of assimilations depending on manner and place of articulation of C_1 in lexical and function words. For both function and lexical words, nasals are the most frequent to assimilate (350 out of 393—89.1%). Overall, stops assimilate in 28 (7.1%) cases and fri-

TABLE II. C_1 Targets and C_2 Triggers for all assimilated function words. The lightly shaded cells highlight assimilations. (a) Function words ending in a [labial]. (b) Function words ending in a [coronal]. (c) Function words ending in a [dorsal].

(a) Function words ending in a [LABIAL]

Assimilation	LABIAL>CORONAL						LABIAL>DORSAL	
C_1 Target / C_2 Triggers PLACE [LABIAL]		n	t,d	ts	z	ʃ		k,g
/m/ 27/583 4.6%	m>n 4.3%	1/76	13/204	4/79	5/177	0/4	m>ŋ 9.3%	4/43
/p, b, f, v/ 0/141	p>t	0/18	0/81	0/10	0/32	0/0	p>k	0/0
Sum 27/724 3.7%	23 / 681 3.4%						4 / 43 9.3%	

(b) Function words ending in a [CORONAL]

Assimilation	CORONAL > LABIAL					CORONAL > DORSAL	
C_1 Target / C_2 Triggers PLACE [CORONAL]		m	p,b	pf	f,v		k,g
/n/ 225 / 1230 18.3%	n>m 16.2%	44 / 187	33 / 142	3 / 4	88 / 703	n>ŋ 29.4%	57 / 194
/t, d/ 4 / 200 2.0%	t>p 2.3%	4 / 43	0 / 21	0 / 2	0 / 107	t>k	0 / 27
/s/ 1 / 1021 0.1%	s>f 0.1%	0 / 138	0 / 161	0 / 1	1 / 534	s>x	0 / 187
/ç/ 2 / 510 0.4%	ç>f 0.6%	0 / 91	0 / 56	0 / 0	2 / 186	ç>x	0 / 177
Sum 232 / 2961 7.8%	175 / 2376 7.4%					57 / 585 9.7%	

(c) Function words ending in a [DORSAL]

Assimilation	DORSAL > LABIAL					DORSAL > CORONAL					
C_1 Target / C_2 Triggers PLACE [DORSAL]		m	p,b	pf	f,v		n	t,d	ts	z	ʃ
/x/ 7 / 459 1.5%	x>f 2.8%	0/70	0/49	0/5	6/94	x>s 0.4%	0/51	0/98	0/24	1/37	0/31
Sum 7 / 459 1.5%	6 / 218 2.8%					1 / 241 0.4%					

TABLE III. C_1 target and C_2 triggers of all assimilated lexical words. The lightly shaded cells highlight assimilations. (a) Lexical words ending in [labial]. (b) Lexical words ending in [coronal]. (c) Lexical words ending in [dorsal].

(a) Lexical words ending in [LABIAL]

Assimilation	LABIAL>CORONAL						LABIAL>DORSAL	
C_2 Trigger / C_1 Target PLACE [LABIAL]		n	t,d	ts	z	ʃ		k,g
/m/ 1 / 34 2.9%	m>n 3.3%	0/5	1/20	0/2	0/2	0/1		0/4
/p, b, f, v/ 0 / 42		0/5	0/26	0/0	0/3	0/2		0/6
Sum 1 / 76 1.3%	1 / 66 1.5%						0 / 10 0%	

(b): Lexical words ending in [CORONAL]

Assimilation	CORONAL>LABIAL					CORONAL>DORSAL	
C_2 Trigger / C_1 Target PLACE [CORONAL]		m	p,b	pf	f,v		k,g
/n/ 97 / 1050 9.2%	n>m 9.4%	14/160	29/285	0/0	46/497	n>ŋ 7.4%	8/108
/t, d/ 24 / 531 4.5%	t>p 4.9%	18/146	3/143	0/1	2/184	t>k 1.8%	1/57
/s, ç, ʃ/ 0 / 386		0/68	0/59	0/0	0/165		0/97
Sum 121 / 1970 6.1%	112 / 1708 6.6%					9 / 262 3.4%	

(c): Lexical words ending in [DORSAL]

Assimilation	DORSAL>LABIAL				DORSAL>CORONAL					
C_2 Triggers / C_1 Target PLACE [DORSAL]		m	p,b	f,v		n	t,d	ts,tʃ	z	ʃ
/ŋ, k, g/ 0 / 342		0/25	0/12	0/78		0/27	0/158	0/13	0/22	0/7
/x/ 5 / 528 0.9%	x>f 2.5%	0/32	0/37	3/53	x>s 0.5%	0/15	0/343	0/18	2/26	0/4
Sum 5 / 870 0.6%	2 / 237 1.3%				2 / 633 0.3%					

catives in 15 instances (3.8%). Of a total of 393 assimilated targets, overwhelmingly the [coronal] sounds (353 out of 393—89.8%) assimilate to the place of a following segment across word boundaries, whereas [labial] (7.1%) and [dorsal] (3.1%) segments usually do not. In general, coronal targets (C_1) by far outnumber the other places of articulation (4931 or 69.8%). The fewest number of targets are [labial] sounds

TABLE IV. Assimilation of function and lexical words combined.

C_1 target				C_2 trigger					
Place	Total	Assimilated		[Labial]		[Coronal]		[Dorsal]	
[Labial]	800	28	3.5%	—		24/747	3.2%	4/53	7.5%
[Coronal]	4931	353	7.2%	287/4084	7.0%	—		66/847	7.8%
[Dorsal]	1329	12	0.9%	9/455	2.0%	3/874	0.3%	—	
Sum	7060	393	5.6%	296/4539	6.6%	27/1621	1.7%	70/900	7.8%

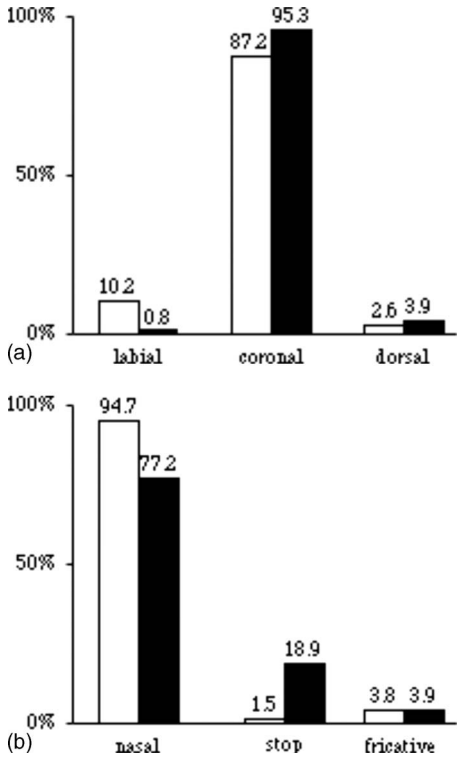


FIG. 1. Relative percentages of regressive place assimilations (based on the total number of assimilated sequences) across different place (a) and manner of articulation (b). Function words are represented by light bars, lexical words by dark bars.

(800 or 11.3%). The only [dorsal] segment—both in function words as well as lexical words—that assimilates is [x].

Note that the analysis did not differentiate between C_1 and C_2 sequences that were within one phrase or sequences that crossed phrase boundaries. Out of the 7060 items analyzed in the data, there were 1174 (16.6%) crossing either a period, a question mark, or a comma in the transcription. Of all 18 cases where C_1 and C_2 were separated by a question boundary, none showed assimilation. Concerning periods, there were overall 310 sequences in this category. There was one (out of 188) assimilation occurring in a [coronal]-[labial] context. Overall a comma separated 848 of the 1174 sequences. In this category, there were 13 assimilations. 10 (out of 319—3.1%) occurred in a [coronal]-[labial] context, 2 (out of 42—2.4%) showed an assimilation of [x] to [f] in front of [f], and two cases (out of 441 possible sequences—0.5%) had an assimilation of [x] to [s] in front of [z]. Thus, although phrase boundaries do impede assimilation, at least for commas, there are cases where assimilation even occurs across those boundaries.

To summarize, across word place assimilations in German is controversial. Some authors claim that such assimilations do not occur (cf. Wurzel, 1970; Vater, 1979; Wiese, 1996), while others assert the opposite (cf. Kohler, 1995). This controversy led us to systematically analyze assimilations across word boundaries in conversational German. The Kiel corpus data suggests that although such assimilations are not frequent, they do occur—overall, approximately 6% of possible assimilatory sequences did undergo a change in place of articulation. We analyzed function and lexical words separately since they are claimed to be different, and indeed, we found a significant difference in the number of assimilations between the two categories although the assimilation patterns were the same. Function words are more likely to assimilate than lexical words. Moreover, the data revealed clear asymmetries in the pattern of assimilations that actually occurred. One asymmetry concerns the place of articulation of the targets (C_1) undergoing assimilation: [coronal] sounds are more frequently assimilated than [dorsal] and [labial] consonants. A second asymmetry is that nasals assimilate more often than stops or fricatives.

The assimilation data we have analyzed and presented are based on the transcriptions of trained phoneticians who noted sequences where assimilation had occurred despite the fact that they had the orthographic as well as the canonical phonetic transcription that could have biased them to perceive the canonical sounds. Speech perception research suggests, however, that although some sounds might seem to be assimilated, there may still be residual cues for listeners to identify the underlying segments (cf. Gow, 2002). We therefore turn to the perception of naive listeners and compare them to the transcriptions of the trained phoneticians.

IV. IDENTIFICATION OF TRANSCRIBED PLACE OF ARTICULATION

The question we investigate next is: Do naive listeners (naive both with respect to the goal of the experiment as well as not having additional information from the context) perceive the assimilated and unassimilated segments from the Kiel dialogues in the same way as trained phoneticians who used speech analyses tools? We conducted two phoneme identification experiments—a forced choice and a free choice task—using material from the Kiel corpus. We opted to focus on nasals (/n/ and /m/) since the choice of assimilated segments was larger than for oral stops and we were able to take stimuli from several speakers thereby lessening speaker dependence (for details see Sec. IV A 1 below).

A. Experiment 1: Phoneme identification

A timed forced-choice identification task was chosen for the first experiment. Subjects had to decide whether the auditory stimuli included either a labial [m] or a coronal [n]. This method was chosen to determine the speed as well as the accuracy of the subjects' decision. Other studies have shown that assimilations can be only partial and that listeners are sensitive to residual cues left (cf. Nolan, 1992; Manuel, 1995; Gow, 2002). Manuel (1995), for example, found that in a sequence [nð] in *win those*, where the /ð/ became a nasal, the place of articulation was not that of a "real" [n], suggesting that some featural information was still available to the listener. Our focus was not just on the assimilated stimuli, but also stimuli that had been labeled as unchanged from the canonical—that is underlying /n/ or /m/ which were spoken and heard as [n] and [m]. The issue was whether the responses to the unchanged stimuli differed across varying contexts—vowel, labial, dorsal, coronal. The crucial conditions with a set of examples are listed in Table V. The segmental context from which the stimuli were extracted is

TABLE V. Examples of stimuli with the vowel [e:] for experiments 1 and 2. Column 1 gives the Kiel transcription. Column 2 provides the orthographic contexts from which the stimuli were extracted and column 3 gives the three conditions—unchanged unassimilated-/m/ and unassimilated-/n/, and assimilated.

Kiel corpus transcription	Example stimuli in orthography	Condition
		Unassimilated-/m/
[e:m]	…von d**em a**chtzehnten Juni?…	Vowel context /m/-vowel
[e:m]	…mit d**em B**ericht…	Labial context /m/-labial
[e:m]	…dann d**em D**ienstag…	Coronal context /m/-coronal
[e:m]	…und d**em g**anzen Kram…	Dorsal context /m/-dorsal
		Unassimilated-/n/
[e:n]	…Freitag, d**en e**rsten…	Vowel context /n/-vowel
[e:n]	…für d**en B**ericht…	Labial context /n/-labial
[e:n]	…in d**en d**eutschen…	Coronal context /n/-coronal
[e:n]	…d**en g**anzen Tag…	Dorsal context /n/-dorsal
		Assimilated
[e:m]	…über d**en B**ericht…	Labial context

double underlined. Since the coronal nasal assimilated most frequently, we only used conditions where /n/ was assimilated to [m].

Our predictions are the following. Most descriptions of assimilations suggest that coronal consonants are more vulnerable to variation in the context of consonants with other places of articulation (cf. Paradis and Prunet, 1991). Consequently, one could expect that labial and dorsal C_2 contexts would leave more acoustic traces in unassimilated-/n/ stimuli than coronal and dorsal C_2 segments influence unassimilated-/m/ stimuli. This would make it more difficult for listeners to come to a definite decision for the unassimilated-/n/ stimuli. Therefore, we expect slower reaction times (RTs) for unassimilated-/n/ in labial and dorsal contexts but no reaction time differences for those items in the vowel or homorganic consonantal context. Insofar as the difference between assimilated-/n/s and unassimilated-/m/s are concerned, we expect no difference in the speed of reaction, assuming that the assimilated-/n/s exhibit complete neutralization. However, whether the assimilated-/n/ items were equally well heard as [m] as the unassimilated-/m/s depends on whether the assimilation as perceived by the transcribers was reasonably complete. Thus, both the reaction time measures as well as percentage of [m] and [n] responses are vital.

1. Materials

The stimuli for the perception task consisted of a vowel-nasal (VN) sequence extracted from real words (CVN or VN), and were taken from 27 different speakers (13 female, 14 male) of the Kiel corpus. At most five items were taken from any given speaker. We thereby kept the segmental context as similar as possible and at the same time were able to make the perception task speaker-independent. The two vowels in the VN sequences we chose were transcribed as either a mid [eː] or a low [ɑ] vowel. The extracted sequences with [ɑ] form possible words: *an* [ɑn] "on, at.acc" and *am* [ɑm] "at.dat," whereas the [eːn] and [eːm] sequences do not. A set of sentences from which the [eː] sequences were extracted is given in Table V and corresponding [ɑm/ɑn] sequences are given in Appendix B.

We cut the VN-items at zero-crossings in order to avoid clicks at item boundaries using both visual as well as auditory information. The first identifiable glottal period was taken as the beginning of the vowel. However, when there was an extensive amount of coarticulation from the preceding segment (i.e., at the word onset), we cut off up to four glottal periods to ensure that the consonantal onset could no longer be perceived. The end of the nasal in the VN-items was determined when the amplitude of the waveform dropped markedly or at the beginning of the closure of the following consonant. Thus, the nasal itself was left untouched, but any contextual information in the following closure would have been removed.

For each vowel (i.e., [eː]/[ɑ]), we chose ten [coronal]#[labial] assimilated sequences (assimilated category), and 10 each of unassimilated [coronal] (unassimilated-/n/) and [labial] (unassimilated-/m/) items. This added up to 60 different stimuli. The unassimilated items were cut out of different contexts (see Table V and Appendix B); three preceded a [labial] consonant, three a [coronal] consonant, two a [dorsal] consonant, and two were originally followed by a vowel. The amplitude of the items was equalized.

2. Subjects and procedure

Overall, 18 undergraduates from the University of Konstanz with no reported hearing disorders served as subjects in the experiment and were paid for their participation. They were tested in groups of 5 or less and were given oral as well as written instructions. A push-button box with two buttons labeled [m] and [n] was placed in front of each subject. They were instructed to listen to the syllables presented over headphones and decide as quickly as possible whether the consonant was [m] or [n] and press the appropriate button with the index finger of their dominant hand. Before the test began,

TABLE VI. Least square means of reaction times for three main categories in all contexts for both [m] and [n] responses [% values are computed for each row by $N_x/(N_{\text{Response[m]}}+N_{\text{Response[n]}})$].

	Response [m]			Response [n]		
Context	*N*	%	RT ms	*N*	%	RT ms
Unassimilated-/m/	1643	93.2	536.3	120	6.8	580.4
/m/-labial	467	89.1	535.1	57	10.9	518.3
/m/-coronal	523	97.9	531.0	11	2.1	573.2
/m/-dorsal	310	88.6	547.9	40	11.4	647.8
/m/-vowel	343	96.6	531.3	12	3.4	582.1
Unassimilated-/n/	405	23.1	547.1	1346	76.9	547.2
/n/-labial	141	26.8	592.9	385	73.2	553.4
/n/-coronal	92	17.6	520.6	432	82.4	528.4
/n/-dorsal	95	27.2	536.2	254	72.8	570.1
/n/-vowel	77	21.9	538.6	275	78.1	536.8
Assimilated (labial context)	1534	87.5	545.8	219	12.5	580.0

the subjects familiarized themselves to the task with practice items, but were given no feedback about the "congruency" of their decisions.[12]

Each item occurred five times during the experiment, adding up to 300 items presented in a randomized order. The sequence of presentation was as follows. Each item was preceded by a warning tone of 300 ms followed by 200 ms of silence. After each test stimulus, there was a pause of 1500 ms where subjects had time to push the button and the next sequence began. Reaction time measurements began at the onset of the nasal segment. The stimuli were played from a SONY DAT recorder and presented over headphones (Sennheiser HD520II). In the setup, a central experimental hardware box connected the DAT recorder, the response boxes and a Macintosh computer, where the reaction times were recorded (Reetz and Kleinmann, 2003). A single experimental session lasted approximately 18 min excluding the practice items.

3. Results

The responses of all 18 subjects went into the reaction time analysis.[13] Responses faster than 200 ms and slower than 1000 ms were disregarded leading to the exclusion of 133 responses (2.5% of the data). None of the subjects showed an exceptionally high number of responses which were too slow or too fast. Reaction times as a dependent variable and the factors *subject* (as random), *response* ([m] or [n]),[14] *underlying* (unassimilated-/m/, unassimilated-/n/, assimilated), *context* (nested under *underlying*) (/n/-coronal, /n/-labial, /n/-dorsal, /n/-V, /m/-coronal, /m/-labial, /m/-dorsal, /m/-V, assimilated), *item* (nested under *underlying* and *context*), *response*×*context* (nested under underlying) and *underlying*×*response* as independent variables were entered into an analysis of variance (ANOVA) with REML estimation.[15] There was a main effect of *context* [$F(65\,181)=9.03$, $p<0.001$] and *response* [$F(15\,181)=15.37$, $p<0.001$], and the interaction of *context*×*response* was also significant [$F(65\,181)=4.70$, $p<0.001$]. *Speaker* and *repetition* were not significant factors in the ANOVA. They are therefore not reported. The least square means of the RT measures for both [m] and [n] responses for each *context* are given in Table VI.

Several pairwise posthoc comparisons were made for the critical conditions, the interpretations of which are summarized below with individual figures.

(i) Recall that based on the analysis of the Kiel corpora transcriptions, the expected congruent responses are [m], for the unassimilated-/m/ category and [n] for the unassimilated-/n/ category. The percentage congruent responses are revealing. For unassimilated-/m/ stimuli, 93% of the responses were [m], and only 7% were [n]. In contrast, for unassimilated-/n/ items, almost a quarter of the stimuli were identified as the opposite [m]—77% [n] versus 23% [m]. Obviously, listeners had more difficulty with the unassimilated-/n/ stimuli than with unassimilated-/m/ items. A *Chi-Square* analysis revealed a significant difference ($\chi^2=1773.63$, $p<0.001$). The reaction times also reflect the same pattern. If we consider the congruent responses, [m] for unassimilated-/m/ and [n] for the unassimilated-/n/, the reaction times across these categories (536 and 547 ms, respectively) are statistically different ($t=2.15, p<0.05$). There is a much larger difference between the reaction times for [m]- and [n]-responses to the unassimilated-/m/ stimuli (536 ms versus 580 ms, $t=2.97$, $p<0.05$). Likewise, there is a significant difference between the incongruent [m]-responses of unassimilated-/n/ and the [n]-responses of unassimilated-/m/ (547 ms versus 580 ms, $t=2.04$, $p<0.05$). The RT of [m] or [n] responses to the unassimilated-/n/ category are essentially identical. This suggests that it was more difficult for the listeners, and hence, they were slower, to give [n] responses to unassimilated-/m/ stimuli when they were uncertain.

(ii) Since there were four contexts, the next point to address is if any particular context is responsible for the worse identification of unassimilated-/n/ than

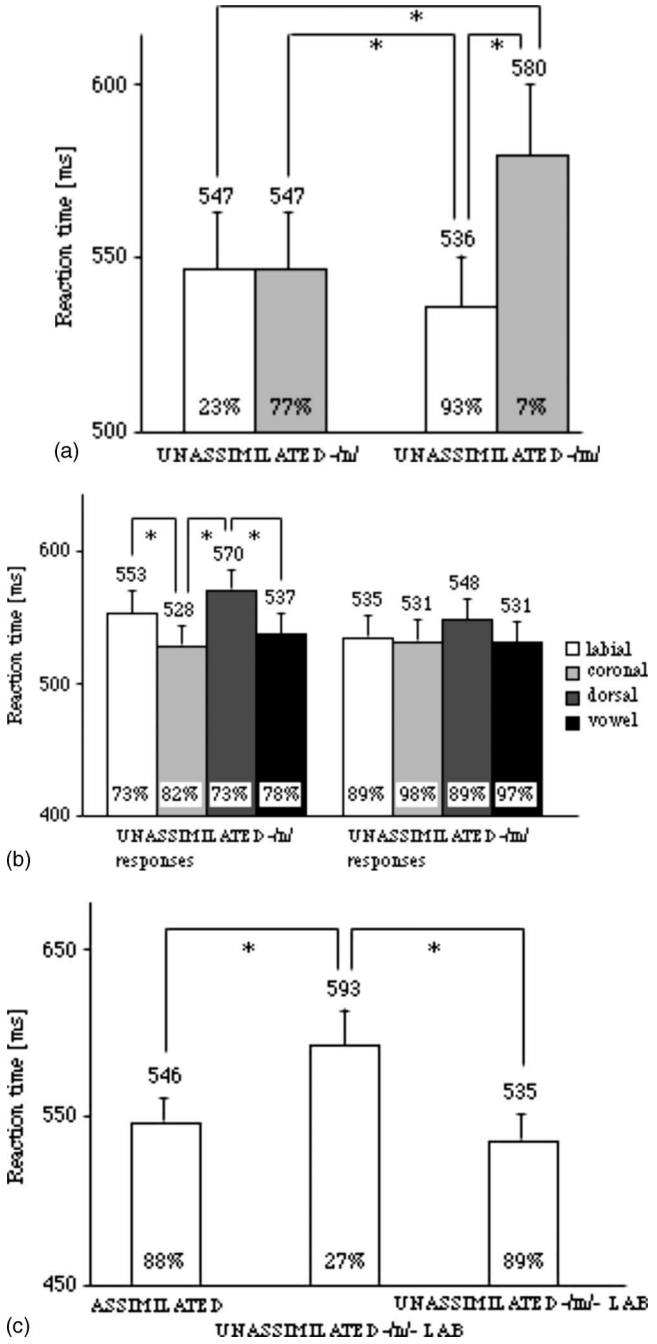


FIG. 2. (a) [m] and [n] responses to unassimilated-/n/ and unassimilated-/m/ stimuli in percent and with their reaction times as bars. Asterisks indicate significant differences in reaction times. White bars represent [m]-responses and gray bars show [n]-responses. (b) [n] responses to unassimilated-/n/ and [m] responses to unassimilated-/m/ stimuli differentiated by context. Percent of responses are given in numbers and the bars represent the reaction times with significant differences indicated by asterisks. (c) Percentages and reaction times for [m] responses to assimilated, unassimilated-/n/-labial, and unassimilated-/m/-labial stimuli. Asterisks indicate significant reaction time differences.

unassimilated-/m/ [see Fig. 2(b)]. With respect to percentage congruent responses, in all contexts more than 89% of the unassimilated-/m/ stimuli were congruently responded to as [m]. This was not so for the unassimilated-/n/ stimuli, where 27% of the responses were [m] in the labial and dorsal contexts. When an unassimilated-/n/ item was preceding another coronal or a vowel, the responses were more comparable to the unassimilated-/m/ stimuli, viz. around 80% [n] responses. To test whether parallel results are reflected in the reaction times, we ran pairwise comparisons across all four contexts—vowel, coronal, dorsal, labial [see Fig. 2(b)]. For the [m] responses to unassimilated-/m/, there were no significant differences in reaction across any of the contexts.

Thus, unassimilated-/m/ (extracted from *dem, am*, etc.) stimuli were heard and reacted to as [m] equally fast regardless of which context they had been extracted from. Would we find the same pattern for [n] responses to the unassimilated-/n/ category stimuli? Based in the corpus analysis, we know that /n/ is more vulnerable to coarticulation from following consonants with different places of articulation. There could therefore be a difference between the contexts dorsal, labial, on the one hand, versus coronal and vowel, on the other. In the former contexts, the /n/ may have more coarticulation cues of the place of articulation of the following dorsal or labial consonant, making it more difficult to label the unassimilated-/n/ as [n] in a reaction time task, whereas in the coronal context, the /n/ is in its ideal environment. The pairwise comparisons confirmed this prediction. The [n] responses to unassimilated-/n/ in coronal context differed significantly from the responses to unassimilated-/n/ in labial context ($t=-2.82$, $p<0.005$) as well as from the dorsal contexts ($t=-3.99$, $p<0.001$). Another significant difference emerged in the comparison of the [n]-responses to unassimilated-/n/ in the dorsal and the vowel contexts ($t=-2.91$, $p<0.005$). There were no further significant differences between any other contexts for the [n]-responses. Thus, the [n]-responses to unassimilated-/n/ in the coronal and vowel contexts, which are the most neutral contexts in terms of coarticulation, are significantly different from the labial and dorsal contexts, We can therefore conclude that the coarticulation cues from the (deleted) following labial and dorsal consonants were strong enough to slow down the subjects' [n] responses to these stimuli. Recall that these consonants had been labeled as [n] by phoneticians who had recourse to both visual and auditory cues and were under no time pressure.

In sum, the labial and dorsal contexts had a slowing down effect on the [n] responses for unassimilated-/n/ stimuli as compared to its homorganic coronal context. This effect is not observed for the unassimilated-/m/ stimuli in the coronal and dorsal contexts in comparison to its homorganic labial context. For the unassimilated-/m/ stimuli, the subjects' speed and their response were unaffected by the context of other places of articulation, from which we may deduce that there were less coarticulation cues which could confuse them. Thus, there was an asymmetry in the stimuli even where trained phoneticians had transcribed the sounds carefully.

(iii) The assimilated stimuli were always (by definition) extracted from a labial context. The crucial question

to gain further insight in how far the assimilations were produced completely is whether these stimuli differ from the unassimilated-/m/ stimuli in the same context. The unassimilated-/m/ stimuli in labial context can be seen as the most prototypically produced labial features without coarticulation and they are taken as clear examples of [m]. Since we found an effect of coarticulation of the labial context in the unassimilated-/n/ stimuli, we also take these for comparison. With respect to percent congruent responses, the [m] responses to the assimilated stimuli and the unassimilated-/m/-labial were almost identical—88% versus 89%. Further, there were no significant differences in reaction times in the [m] or [n] responses to these categories. From this we can conclude that subjects were equally fast in responding to the assimilated [m] and the canonical /m/ stimuli (e.g., [e:m] from *über de**n** **B**ericht* versus [e:m] from *mit de**m** **B**ericht*).

As for the reaction time of [m] responses to unassimilated-/n/-labial stimuli, these were different from the [m] responses to the other two categories (assimilated versus unassimilated-/n/-labial $t=-4.08$, $p<0.001$; unassimilated-/n/-labial versus unassimilated-/m/-labial $t=-4.64$, $p<0.001$), indicating that although there was sufficient coarticulation, these stimuli were different from those that were considered by the transcribers as real assimilated or canonical unassimilated-/m/ items. Crucially, there is no difference between the [m] responses in the assimilated and the unassimilated-/m/-labial categories ($t=-1.65$, $p<0.1$). Thus, for listeners, the assimilated stimuli were similar to the unassimilated-/m/-labial but not to the unassimilated-/n/-labial.

Recall that the task in experiment 1 was forced choice where subjects had to choose between [m] or [n] as possible responses. To determine in how far the forced choice task of experiment 1 created a possible bias in the subjects' responses, we ran a second experiment where the listeners were free to choose and write down what they heard.[16] Our hypothesis was that if we obtained the same pattern of results, then we could conclude that the context-dependent responses of unassimilated-/n/ stimuli was caused by the fact that the listeners were forced to choose between [m] or [n]. Further, we also wanted to examine the pattern of responses to the dorsal stimuli since in experiment 1 the listeners had no option of providing dorsal responses.

B. Experiment 2: Phoneme transcription task

1. Material and design

In order to analyze if possible confounds in the forced-choice task could have affected the results, we used a phoneme-identification task where subjects could write down in a booklet what fragment they heard. The stimuli were identical to experiment 1 except that there was a longer pause between two items (2500 ms instead of 1500 ms), sufficient for writing the syllables but not too much time to think about the stimuli. Each page in the booklet had space for ten items. Warning tones were added after every ten items, prompting subjects to turn to the next page of the booklet. This was done to ensure that if a subject missed an item, it was possible to correctly resume at the beginning of the next page. Thus, as in experiment 1, subjects listened to 300 stimuli.

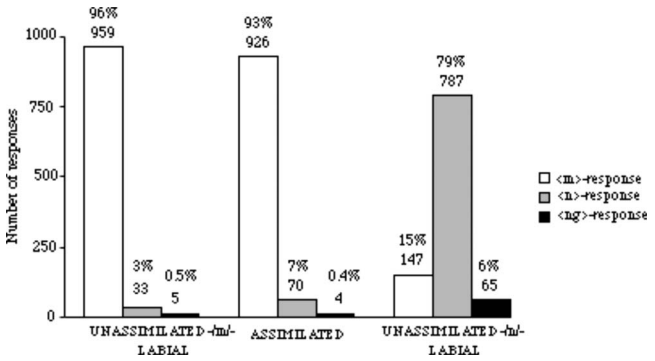


FIG. 3. Total number of responses and percentages within the three main categories.

2. Subjects and procedure

Ten students from the University of Konstanz served as subjects, and none had taken part in the earlier experiment. They were tested individually and were paid for their participation. The set up and equipment was the same as in experiment 1. Written instructions were given to the subjects prior to the experiment and they received the same practice items as before. They were asked to write down what they heard as quickly and accurately as possible. No instruction was given with reference to nasals, syllables or the "wordness" of the items. Given German orthography, if subjects heard nasals, we expected subjects to transcribe them using one of the three possible responses ⟨m⟩, ⟨n⟩, or ⟨ng⟩.

3. Results

In all, there was only one missing response and three were not a nasal. These four items were discarded (0.13%). The nasal responses were split up into the three main categories as above (labial, assimilated, and coronal), based on the original labeling in the Kiel corpus. A total of 2996 transcribed items went into the analysis. Across all categories subjects heard 2032 ⟨m⟩ (67.8%), 890 ⟨n⟩ (29.7%), and only 74 ⟨ng⟩ (2.5%), of which 41 (i.e., 55.4%) come from unassimilated-/n/ in a dorsal context.

Within the individual categories, the nasal segments were transcribed as follows (see Fig. 3). Unassimilated-/m/ segments were transcribed as ⟨m⟩ in 959 cases (96.2%), ⟨n⟩ in 33 instances (3.3%), and ⟨ng⟩ in 5 (0.5%) cases. Assimilated tokens were transcribed as ⟨m⟩ in 926 cases (92.6%), as ⟨n⟩ in 70 (7.0%) cases, as ⟨ng⟩ in four cases (0.4%). Unassimilated-/n/ were transcribed as ⟨n⟩ in 787 (78.8%) cases, ⟨m⟩ 147 times (14.7%), and ⟨ng⟩ in 65 (6.5%) instances.

Insofar as congruent responses are concerned, the percentage of [m] responses to assimilated and unassimilated-/m/ categories is far higher than the corresponding [n] responses to the unassimilated-/n/ category (93%, 96%, versus 79%).

4. Discussion

The free choice task was taken on to ensure that the incongruent responses in experiment 1 were not due to the fact that subjects were forced to choose between two nasals. In particular, we were concerned that the large number of ⟨m⟩ responses to unassimilated-/n/ stimuli was biased by the forced choice task. However, experiment 2 shows that this was not the case. First, there were only three non-nasal responses, and second, 97.6% of the entire responses were transcribed as ⟨m⟩ or ⟨n⟩.

In fact the pattern of results was the same as in experiment 1. On the whole, the unassimilated-/n/ stimuli were more difficult to identify congruently as ⟨n⟩ (79%) and were subject to context dependent responses, as compared to the unassimilated-/m/ or assimilated items, both of which were congruently identified as ⟨m⟩, 96% and 93%, respectively. As in experiment 1, the unassimilated-/n/ stimuli in the context of labial consonants were identified as ⟨m⟩ 15% of the time (experiment 1: 27%). In contrast, there were only 3% ⟨n⟩ responses to unassimilated-/m/ items. Overall, the accuracy of experiment 1 for [labial] and assimilated tokens was even higher in experiment 2, possibly due to the longer time subjects had for their decisions. The results for the assimilated category are very much the same as in experiment 1. They were largely perceived as [labial], indicating the completeness of assimilation. In general, this experiment replicates the same asymmetry we observed already in the identification task and the corpora analysis. One remaining issue is the acoustic differences between the different conditions, critically between the assimilated labial, the canonical unassimilated-/m/ against the unassimilated-/n/-coronal. Since the assimilated nasals did not differ in perception from the canonical unassimilated-/m/, one would conjecture that the acoustic differences would also be minimal.

C. Acoustic measurements

One important issue that has also been reflected in the literature on place assimilation is the question whether acoustic cues can be found that relate to listeners' decisions for [n] or [m] (e.g., Nolan, 1992; Gow, 2002; Dilley and Pitt, 2007). Following Dilley and Pitt's (2007) approach, we investigated the stimuli from the experiments. In their study, they compared assimilated segments with their underlying counterparts. Since their results are based on the variation in the F2 of the preceding vowel, we took the same measure and applied it to items from our perception test. We opted for the analysis of the most prototypical items. Therefore, we compared the assimilated items with unassimilated-/m/ stimuli in labial context and unassimilated-/n/ stimuli in coronal context. Since the number of items from the experiments was too small for calculating an ANOVA, we randomly selected additional items from the Kiel corpus.

There is one important difference between our stimuli compared to Dilley and Pitt's (2007): In our stimuli, the final consonant, in our case the nasal, was not deleted and acoustic information on place of articulation can also be extracted from the nasal segment, therefore, we also took F2 measurements at the midpoint of the nasal segment.

1. Method

We measured the difference in the F2 frequency values in hertz between the middle of the vowel and immediately before the beginning of the nasal murmur of all 20 assimilated items, six unassimilated-/m/-labial, and six unassimilated-/n/-coronal items that were used in the perception studies as an indication for the amount of possible assimilation. In order to base a statistical analysis on a more thorough database, we randomly added the measurements of four assimilated and 18 unassimilated-/m/-labial and 18 unassimilated-/n/-coronal items with the respective vowels. Overall, the measurements of 72 items were analyzed—36 for each vowel (i.e., [eː]/[ɑ]), 24 for each condition (i.e., assimilated, unassimilated-/m/-labial, and unassimilated-/n/-coronal). As in Dilley and Pitt, a mixture of automatic and hand taken measurements was performed (Dilley and Pitt, 2007). Formant values were taken form the estimation provided by PRAAT (Vers. 4.6.04, Boersma & Weening) and wide-band spectrograms. In case that the estimated formant values differed from the spectrograms we followed the spectrograms readings. Dilley and Pitt could measure only the difference between midpoint and endpoint of vowels to gain information about the place of articulation of the upcoming segments, since their items included cases where the consonant in question had been deleted. Since the nasal consonant was never deleted in our case, we were able to measure the F2 frequency in the midpoint of the nasal segments (F2 measurements on the nasal differ for labial and coronal nasal consonants, cf. Stevens, 1998, pp. 487–507). The F2 values at the midpoint of the nasals were measured the same way as in the vowels.

2. Results

F2 differences in the midpoint and endpoint of preceding vowels were subject to an ANOVA with *condition* (assimilated, unassimilated-/m/-labial, and unassimilated-/n/-coronal) and *vowel* as independent variable, as well as the interaction of the two factors (*vowel*×*condition*). Post-hoc tests were performed for the contrasts between the *conditions*. Figure 4(a) summarizes the results for the F2 differences.

As for the F2 difference analysis, there was a main effect of both *condition* [$F(2,66)=10.7106$, $p<0.002$] and *vowel* [$F(1,66)=3.3052$, $p<0.05$], but no significant interaction. A post hoc test revealed that unassimilated-/n/-coronal items were significantly different from assimilated ($t=-2.317$, $p<0.05$) and unassimilated-/m/-labial ($t=2.1242$, $p<0.05$) items, but the latter two were not significantly different from each other.

For the F2 measurements taken at the midpoint of the nasal consonants [see Fig. 4(b)] we used the same ANOVA design and gained the following results: There was a main effect of *condition* [$F(2,66)=5.1775$, $p<0.01$], but no effect of *vowel*, and no interaction. A post-hoc test showed that

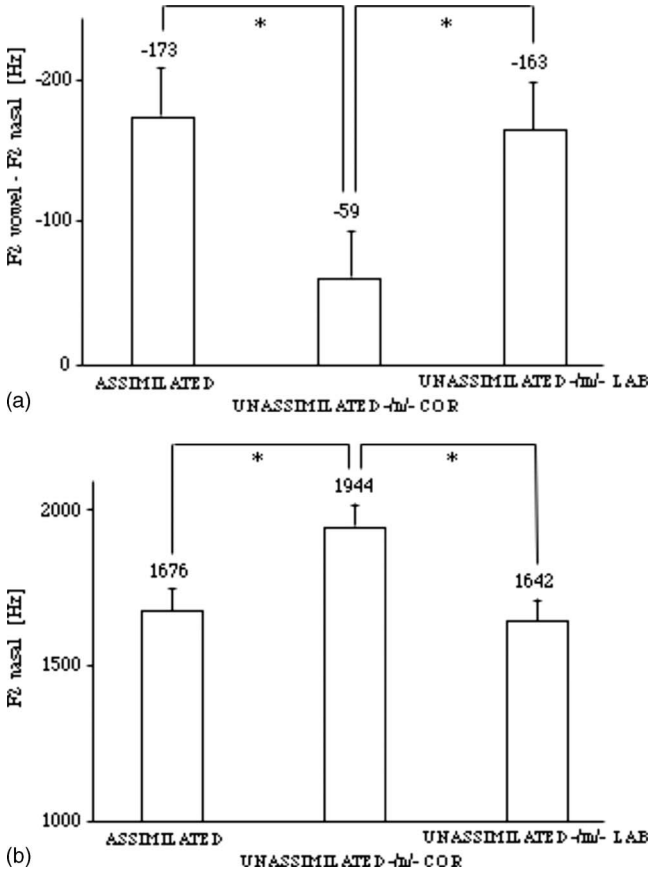


FIG. 4. (a) Differences between F2-frequency measures at the middle and at the end of the vowel in hertz. Significant differences between these difference values are marked by asterisks. (b) Least square means of F2-frequencies at nasal midpoints for the investigated conditions, significant differences are indicated with asterisks.

unassimilated-/n/-coronal items were significantly different from assimilated ($t=-2.605$, $p<0.02$) and unassimilated-/m/-labial ($t=2.9385$, $p<0.005$) items, but the latter two were not significantly different from each other. Figure 4(b) depicts the least square means of the nasal F2 measurements.

3. Discussion

The F2 differences between vowel and nasals and the F2 measurements on the nasals correspond to our perception results. There is no significant difference between the F2 of the assimilated coronals and the canonical unassimilated-/m/. Nor does the nasal F2 differ in these two categories. Corresponding to the perception results, there is a significant difference both in the F2 of the nasal and the F2 difference for the assimilated and unassimilated-/m/-labial nasals on the one hand and the unassimilated-/n/-coronal on the other. The results indicate that subjects take these acoustic cues as basis for their decision when deciding on whether they heard [m] or [n].

V. GENERAL DISCUSSION

The focus of this paper was to investigate the extent to which regressive place assimilations across words exist in conversational German and how listeners perceive them. Analyzing the Kiel corpus of spontaneous speech (IPDS, 1994; 42 speakers, approximately 4 h of recordings) we found that place assimilation does occur across words in German in approximately 6% of all possible sequences of consonants differing in place of articulation (cf. Sec. III C). This is slightly less than the percentage of assimilation reported by Dilley and Pitt (2007) for American English. We based our conclusions on the transcriptions made by trained phoneticians who had recourse to the speech signal as well as the context. Function words were more likely to assimilate than lexical words. Moreover, there was an asymmetry in the direction of assimilation. According to the transcriptions, coronal sounds [t, s, ʃ, ç, n] were more likely to assimilate than labial [p, f, m] or dorsals [k, x, ŋ] (see Table IV). A further asymmetry concerned the manner of articulation of consonants which were more likely to assimilate; nasal consonants were far more likely to assimilate than obstruents.

Exploiting the asymmetries, we ran two perception experiments (forced choice and free choice) to test how fast and accurately naive listeners' responses would correlate with the transcriptions. The two perception experiments using stimuli (from 27 speakers) labeled in the corpora as assimilated (/n/>[m]) or unassimilated (unassimilated-/n/, unassimilated-/m/), showed that the Kiel transcription is very accurate concerning regressive place assimilations. There is a high correlation between the phoneticians' transcriptions and the listeners' judgments. Thus, when the subjects had to decide whether they heard a [coronal] or a [labial] nasal in a VN-syllable without additional information from context, they were very accurate for the assimilated and the unassimilated-/m/ stimuli. The assimilated sounds were overwhelmingly transcribed as [m] both in a speeded forced choice task ([m] or [n]) as well as in a free identification task. In the identical labial context, the reaction times for unassimilated-/m/ and assimilated also did not differ [see Fig. 2(c)]. That is, subjects were equally fast in making an [m] response to stimuli which were really /m/ and those that had been labeled as assimilated from /n/ to [m] in a labial context.

The only nonequivalence between the transcriptions and the listeners' responses was the unassimilated-/n/ category in the context of labial or dorsal segments. If at all, the transcribers were conservative in their judgments of which sounds assimilated as is indicated by the high amount of variation in the /n/-category. As compared to unassimilated-/m/ which was congruently identified as [m] 93% of the time in experiment 1, only 77% of the unassimilated-/n/ stimuli were identified as [n]. The same pattern shows in experiment 2: 79% [n] responses to unassimilated-/n/ stimuli as compared to 96% [m] responses to unassimilated-/m/ stimuli and 93% [m] responses to the assimilated stimuli. The reaction times in experiment 1 also correspond to the accuracy data. The differences in the RT of congruent and incongruent responses of unassimilated-/m/ and unassimilated-/n/ responses are revealing. First, the incongruent [n] responses to unassimilated-/m/ stimuli are significantly slower than the corresponding incongruent [m] responses to unassimilated-/n/ stimuli, indicating that the former were less transparent for the listeners. Second, there is a stronger context effect for the unassimilated-/n/ stimuli

than for the unassimilated-/m/ stimuli. For the latter, there were no differences in reaction to the congruent [m] responses. However, the reaction times for the congruent [n] responses to the unassimilated-/n/ stimuli differed by context. For instance, in the neutral vowel and identical coronal contexts, the reaction times were much faster than in the labial and dorsal contexts. One could argue that coarticulation effects of the following deleted segments cause the incongruent responses, but this does not explain the asymmetry between unassimilated-/m/ and unassimilated-/n/ stimuli. That is, the coarticulation effects on the unassimilated-/n/ stimuli were presumably strong enough for listeners to respond to them with [m] as fast as they responded to the unassimilated-/n/ stimuli in their ideal coronal context. Almost a quarter of the unassimilated-/n/ stimuli were incongruently identified as [m] in both experiments. We see this as an indication that the transcribers of the Kiel corpus were "conservative" and labeled the unassimilated-/n/-labial as [n] rather than [m]. Since experiment 2 used a free-choice task, and the results were essentially the same to experiment 1, it is unlikely that subjects were forced to choose ⟨n⟩ as an answer as substitute for any other perceived segment.

We further investigated the acoustic patterns across the stimuli used in the perception experiments (Sec. IV C). Since our stimuli consisted of vowel-plus-nasal syllables, following Dilley and Pitt (2007) we took the F2 measures of the middle and end of the vowel, we also examined F2 at the nasal midpoint. Corresponding to the perception results, we found that the change in the F2 from the middle to the end of the vowel did not significantly differ between the unassimilated-/m/ and assimilated consonants. Similarly, the nasal formant measure did not differ between these categories indicating that the assimilated tokens shared these acoustic categories with the canonical /m/.

Both the perception results and acoustic analysis of the stimuli suggest that segments labeled as assimilated by the transcribers are indeed recognized not as tokens of the underlying words, but as perfect instances of the changed sound, i.e., complete assimilations do occur in running speech (/*n*/ > [*m*] in a labial context). Actually, some assimilated tokens are judged by subjects in our experiments to be [m] 100% of the time. Clearly however, there exists gradience in the assimilation as we have seen in the response to the unassimilated-/n/-labial stimuli. Although transcribers labeled them as [n] they were often perceived as [m]. Gradualness of assimilation is most important for the [coronal]-category where we see the greatest amount of (response) variation.

The asymmetry between coronal versus dorsal and labials both in production analysis (coronal consonants assimilate more than the others) and in perception (coronals vary most in perception) has been frequently noted in the literature (cf. Lahiri and Evers, 1991; Paradis and Prunet, 1991; Ghini, 2001). The unmarkedness and asymmetry of coronals are tackled by markedness rules or other phonological principles (cf. Clements, 2001) or built into a recognition model with underspecification as in FUL (Lahiri and Reetz, 2002).

The results of the corpus study as well as the two experiments are fully compatible with the FUL model of speech perception (Lahiri and Reetz, 2002). The model assumes that segments can be specified with a [labial] or [dorsal] feature for place of articulation, but do not possess a [coronal] feature, that is, they are underspecified. Evidence for underspecification has been presented in different languages and for different phonological processes (among others, Lahiri and Evers, 1991; Lahiri and Marslen-Wilson, 1991; Ghini, 2001, Wheeldon and Waksler, 2004; Scharinger, 2006; cf. also Paradis and Prunet, 1991). Segments specified for features [labial] or [dorsal] lead to corresponding labial and dorsal places of articulation in production. Segments which are underspecified for place of articulation are produced by the default feature [coronal]. Moreover, since these are underlyingly not specified for place of articulation, they can assimilate more easily to the place of articulation of neighboring [labial] and [dorsal] segments, but not *vice versa*. There can be instances where segments are produced in-between two categories due to overlapping gestures. FUL predicts that if this assimilation takes place before the default production rule applies, assimilation can lead to a complete neutralization of the place of articulation contrast. This is what we find in the corpus study, namely assimilations almost exclusively occur with [coronal] segments that assimilate to either [labial] or [dorsal] place of articulation, but not the other way around.

The model can also explain the finding that in general, RTs are slower for coronals than for labials. During recognition, features are extracted from the speech signal and matched directly onto lexical representations. However, for [coronal] sounds, there is no feature in the lexical representation that can be matched. Therefore, a nonmismatch condition arises. On the other hand, [labial] features can be matched onto a labial feature in a lexicon. Although there is no direct claim that nonmismatch conditions are always slower than matching conditions, the results are not unexpected from a theoretical point of view. However, there need to be more studies in order to exactly determine in how far there is a time advantage for matching versus nonmismatching conditions.

This was a first attempt to examine the perceptual consequences of such assimilations in conversational speech and it appears that if a coronal consonant is assimilated to the following consonant, it is perceived as such even when presented without any context. Further studies are necessary to see if this pattern or results hold for other languages as well.

ACKNOWLEDGMENTS

This work was in part supported by the Deutsche Forschungsgesellschaft (SFB 471, SPP 1234, and the Leibniz Preis awarded to A.L.) and the Ministry of Science, Research, and the Arts of Baden-Württemberg. We also wish to thank our anonymous reviewers for their very useful comments and suggestions.

APPENDIX A: ALL PRONUNCIATION VARIATIONS OF *einverstanden* IN THE CORPUS IN IPA TRANSCRIPTION

	Phonetic transcription	Deviations from canonical transcription
	['ʔaɪnfɐˌʃtandən]	Canonical transcription, no deviations
i	['aɪnfɐˌʃtaŋ̍n]	1 segment deletion, 2 glottalizations
ii	['aɪnfɐˌʃtanʔn]	1 segment deletion, 1 glottalization, 1 weakening
iii	['aɪ̰nfɐˌʃtaŋ̍]	2 segment deletions, 2 glottalizations
iv	['aɪ̰nfɐˌʃtan̰]	3 segment deletions, 1 glottalization
v	['aɪnfɐˌʃtandn]	1 segment deletion, 1 glottalization
vi	['aɪnfɐˌʃtanhn]	1 segment deletion, 1 glottalization, 1 weakening
vii	['aɪmfɐˌʃtan̰]	3 segment deletions, 1 glottalization, 1 assimilation
viii	['ʔaɪnfɐˌʃtan̰]	2 segment deletions, 2 glottalizations
ix	['ʔaɪnfɐˌʃtan̰ʔn]	1 segment deletion, 1 glottalization, 1 weakening
x	['ʔaɪnfɐˌʃtan̰]	2 segment deletions, 2 glottalizations
xi	['ʔaɪnfɐˌʃtaŋ̍n]	1 segment deletion, 2 glottalizations
xii	['ʔaɪnfɐˌʃtandn]	1 segment deletion, 1 glottalization
xiii	['ʔaɪntfɐˌʃtanhn]	1 segment deletion, 1 glottalization, 1 weakening
xiv	['ʔaɪnfɐˌʃtanhn]	1 segment deletion, 1 weakening
xv	['aɪnfɐˌʃtann]	2 segment deletions, 1 glottalization
xvi	['aɪntfɐˌʃtan̰]	2 segment deletions, 2 glottalizations, 1 insertion
xvii	['aɪntfɐˌʃtaŋ̍n]	1 segment deletion, 2 glottalizations, 1 insertion
xviii	['aɪnvɐˌʃtann]	2 segment deletions, 1 glottalization, 1 voicing
xix	['ʔaɪmfɐˌʃtanhn]	1 segment deletion, 1 weakening
xx	['aɪnfɐˌʃanʔn]	2 segment deletions, 1 weakening
xxi	['aɪnfɐˌʃtan̰]	3 segment deletions, 1 glottalization
xxii	['aɪnfɐˌʃtan̰]	3 segment deletions, 1 glottalization
xxiii	[nfɐʃtan̰]	4 segment deletions, 1 glottalization

APPENDIX B: EXAMPLES OF CONTEXTS FROM WHICH [ɑm] AND [ɑn] STIMULI WERE EXTRACTED FOR EXPERIMENTS 1 and 2. FOR DETAILS SEE TABLE V

Kiel corpus Transcription	Example stimuli in orthography	Conditions
		Unassimilated-/m/
[ɑm]	…daram **am e̱**rsten…	Vowel context /m/-vowel
[ɑm]	…das denn **am Ḇ**esten	Labial context /m/-labial
[ɑm]	…wir **am s̱**echsten	Coronal context /m/-coronal
[ɑm]	…wir **am g**ünstigsten…	Dorsal context /m/-dorsal
		Phoneme-/n/
[ɑn]	…sieht das d**ann a̱**us…	Vowel context /n/-vowel
[ɑn]	**Dann ḇ**rauchen wir…	Labial context /n/-labial
[ɑn]	Ist d**ann ḏ**er…	Coronal context /n/-coronal
[ɑn]	…aber m**an ḵ**ann…	Dorsal context /n/-dorsal
		Assimilated
[ɑm]	Und d**ann ḇ**rauchen wir…	Labial context

[1]We ignored the "–h" symbol in the Kiel transcriptions since it has many phonetic correlates (e.g., aspiration, release) which are not relevant here. We translated the SAMPA transcription of the Kiel corpus into standard IPA transcription.

[2]We did not treat glottalization as instance of complete deletion, rather as some remnant of a severely reduced segment to keep the two processes apart.

[3]Neutralizations occur when speakers eliminate contrastive featural contrasts of segments in speech production. For instance, when they produce a segment such as /n/—underlyingly [coronal]—as a [labial] [m] due to a complete assimilation to the place of articulation of an upcoming [labial] segment, such as [b]. Complete means that the resulting [m] (underlyingly [n]) is not different from an underlying /m/ being produced as [m].

[4]Regressive assimilation occurs when in a sequence of two segments S_1 and S_2, S_1 assimilates in some feature(s) to S_2. Progressive assimilation occurs, when S_2 assimilates to S_1.

[5]Throughout this paper, we use the following convention for the description of letters and sounds. The sign ⟨ ⟩ is used to refer to orthography, [ ] indicates phonetic transcription, / / is used for underlying segments, and { } encloses morphemes.

[6]Along with Wiese (1996), Benware (1986) sees the "phonological word" as the only domain where regressive place assimilation can occur. He cites Kallmeyer (1981) for a case of regressive place assimilation in *kaputt gegangen* "has broken down," where the final /t/ of *kaputt* "ruined" is pronounced with a [k]. The phrase *kaputt gehen* consisting of two words is interpreted as a single phonological word in the sense that they form a very close unit, different from usual words in a phrase (Benware, 1986, p. 129).

[7]Since final devoicing affects all places of articulation, we did not differentiate between voiced versus voiceless segments.

[8]If there is a deletion and no assimilation on the preceding segment, it is not clear, whether the deleted segment itself was assimilated. If the preceding segment assimilates, it is not clear, whether the deleted segment triggered the assimilation, or the first segment of the upcoming word.

[9]The features are based on Lahiri and Reetz (2002). Palatals are assumed to be [coronal], as in many phonological accounts (e.g., Lahiri and Evers, 1991; Clements and Hume, 1995; Kenstowicz, 1994; for a different view, see for example Hall, 2000). The segments [x, ç] are assumed to be underlyingly placeless since the place of articulation of the preceding vowel determines the place of articulation of the fricative—[coronal] after front vowels, [dorsal] after back vowels. For sake of simplicity, we refer to the underlying fricatives as /x/ or /ç/.

[10]The fricative [x] is the only [dorsal] consonant function words end with. Due to final devoicing, only voiceless obstruents occur in C_1 target position.

[11]Almost all the cases of /m/ assimilating to /n/ could also be analyzed as being a wrong case-marking, a phenomenon that is well known for many German speakers (Bayer and Brandner, 2004; Schiering, 2005); *den* "the-*.accusative*" instead of *dem* "the.*dative*," etc. However, here we treated them as any other case of assimilation.

[12]We use the term congruent for responses where the transcription of the corpus was the same as the subjects' decision and incongruent for the opposite case.

[13]The analysis was carried out using SAS statistic software JMP, version 5.0.1.2.

[14]Since we were interested in the influence of the response on the reaction time, the responses are treated as a factor.

[15]The residual maximum likelihood (REML) estimation does not substitute missing values with estimated means and does not need synthetic denominators; rather the individual factors are tested against the whole model. This method is more conservative than the traditional expected mean squares estimation. Not significant results did not reach the 5% level.

[16]Subjects had only two possible response buttons, i.e., [n] or [m] to choose from in experiment 1. As can be seen, especially unassimilated-/n/ items in labial context produced a high amount of incongruent responses. This is arguably due to coarticulatory cues. For items in dorsal context, one could also expect coarticulatory cues influencing subjects' responses. However, it is not clear, how subjects would react in this situation, since there was no possibility to indicate "something else." In order to examine the nature of incongruent responses further, we opted for a free transcription task, where subjects could write what they heard without being restricted to two responses, in fact without being restricted to a nasal response at all.

Bayer, J., and Brandner, E. (**2004**). "Klitisiertes *zu* im Bairischen and Alemannischen," (Cliticized to in Bavarian and Alemannic) in *Morphologie und Syntax deutscher Dialekte und historische Dialektologie des Deutschen*, edited by F. Patocka and P. Wiesinger (Praesens Edition, Wien), pp. 160–188.

Benware, W. A. (**1986**). *Phonetics and Phonology of Modern German* (Georgetown University Press, Washington, DC).

Bybee, J. (**2007**). *Frequency of Use and the Organization of Language* (Oxford University Press, Oxford).

Clements, G. N. (**2001**). "Representational economy in constraint-based phonology," in *Distinctive Feature Theory*, edited by T. A. Hall (Mouton de Gruyter, Berlin), pp. 71–146.

Clements, G. N., and Hume, E. V. (**1995**). "The internal organization of speech sounds," in *The Handbook of Phonological Theory*, edited by J. A. Goldsmith (Blackwell, Cambridge, MA).
Dilley, L. C., and Pitt, M. A. (**2007**). "A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition," J. Acoust. Soc. Am. **122**, 2340–2353.
Dressler, W., Fasching, P., Chromec, E., Wintersberger, W., Leodolter, R., Stark, H., Groll, G., Reinhart, J., and Pohl, H. D. (**1972**). "Phonological fast speech rules in colloquial Viennese German," Wiener Linguistische Gazette **1**, 1–30.
Ghini, M. (**2001**). *Asymmetries in the Phonology of Miogliola* (Mouton de Gruyter, Berlin).
Gow, D. W., Jr. (**2002**). "Does English coronal place assimilation create lexical ambiguity?," J. Exp. Psychol. Hum. Percept. Perform. **28**, 163–179.
Hall, T. A. (**1992**). *Syllable Structure and Syllable Related Processes in German* (Max Niemeyer Verlag, Tübingen).
Hall, T. A. (**1999**). "Phonotactics and the prosodic structure of German function words," in *Studies on the Phonological Word*, edited by T. A. Hall and U. Kleinhenz (John Benjamins, Amsterdam), pp. 99–131.
Hall, T. A. (**2000**). *Phonologie—eine Einführung* (Phonology—An introduction) (Walter de Gruyter, Berlin).
IPDS—Institut für Phonetik und digitale Sprachverarbeitung (**1994**). *The Kiel Corpus of Spontaneous Speech, CD-ROM I-III* (IPDS, Kiel).
Kabak, B., and Schiering, R. (**2006**). "The phonology and morphology of function word contractions in German," J. Comparative Germanic Linguistics **9**, 53–99.
Kaisse, E. M. (**1985**). *Connected Speech: The Interaction of Syntax and Phonology* (Academic, Orlando).
Kallmeyer, W. (**1981**). "Aushandlung und Bedeutungskonstitution," (Negotiation and constitution of meaning) in *Dialogforschung—Jahrbuch des Instituts für Deutsche Sprache, 1980*, edited by P. Schröder and H. Steger (Pädagogischer Verlag Schwann, Mannheim), pp. 89–127.
Kenstowicz, M. (**1994**). *Phonology in Generative Grammar* (Blackwell, Oxford).
Kohler, K. J. (**1990**). "Segmental reduction in connected speech in German: Phonological facts and phonetic explanations," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht), pp. 69–92.
Kohler, K. J. (**1995**). *Einführung in die Phonetik des Deutschen*, 2nd ed. (Introduction into the phonetics of German) (Erich Schmidt Verlag, Berlin).
Kohler, K. J., Pätzold, M., and Simpson, A. P. (Eds.) (**1995**). "From scenario to segment—the controlled elicitation, transcription, segmentation and labelling of spontaneous speech," *AIPUK* (IPDS, Kiel), Vol. **29**.
Lahiri, A., and Evers, V. (**1991**). "Palatalisation and coronality," in *The special Status of Coronals: Internal and External Evidence*, edited by C. Paradis and J.-F. Prunet (Academic, San Diego), pp. 79–100.
Lahiri, A., and Marslen-Wilson, W. D. (**1991**). "The mental representation of lexical form: A phonological approach to the recognition lexicon," Cognition **38**, 245–294.
Lahiri, A., and Reetz, H. (**2002**). "Underspecified recognition," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 637–675.
Local, J. (**2003**). "Variable domains and variable relevance: interpreting phonetic exponents," J. Phonetics **31**, 321–339.
Manuel, S. Y. (**1995**). "Speakers nasalize /ð/ after /n/, but listeners still hear /ð/," J. Phonetics **23**, 453–476.
Nespor, M., and Vogel, I. (**1986**). *Prosodic phonology* (Foris Publications, Dordrecht/Riverton).
Nolan, F. (**1992**). "The descriptive role of segments," in *Papers in Laboratory Phonology II—Gesture, Segment, Prosody*, edited by G. J. Docherty and R. Ladd (Cambridge University Press, Cambridge), pp. 261–280.
OED (**1989**). *Oxford English Dictionary*, 2nd ed. (Oxford University Press, Oxford).
Ogden, R. A. (**1999**). "A declarative account of strong and weak auxiliaries in English," Phonology **16**, 55–92.
Paradis, C., and Prunet, J.-F. (Eds.) (**1991**). *Phonetics and Phonology 2: The Special Status of Coronals: Internal and External Evidence* (Academic, San Diego).
Phillips, B. S. (**2001**). "Lexical diffusion, lexical frequency, and lexical analysis," in *Frequency and the Emergence of Linguistic Structure*, edited by J. Bybee and P. J. Hopper (John Benjamins, Amsterdam), pp. 123–136.
Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E. (**2006**). "Buckeye Corpus of conversational speech," (1st release), Department of Psychology, Ohio State University: Columbus, OH.
Reetz, H., and Kleinmann, A. (**2003**). "Multi-subject hardware for experiment control and precise reaction time measurement," in Proceedings ICPhS Barcelona, pp. 1489–1492.
Scharinger, M. (**2006**). "The representation of vocalic features in vowel alternations. Phonological, morphological and computational aspects," Ph.D. thesis, Konstanz, Konstanz Online Publication System, http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-24341 (Last accessed on 1/7/2008).
Schiering, R. (**2005**). "Flektierte Präpositionen im Deutschen? Neue Evidenz aus dem Ruhrgebiet," (Inflected prepositions in German? New evidence from the Ruhr area), Zeitschrift für Dialektologie und Linguistik **72**, 52–79.
Selkirk, E. O. (**1984**). *Phonology and Syntax: The Relation Between Sound and Structure* (The MIT Press, Cambridge, MA).
Snoeren, N. D., Hallé, P. A., and Segui, J. (**2006**). "A voice for the voiceless: Production and perception of assimilated stops in French," J. Phonetics **34**, 241–268.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).
Vater, H. (Ed.). (**1979**). *Phonologische Probleme des Deutschen* (Phonological problems of German) (Gunter Narr Verlag, Tübingen).
Wheeldon, L., and Waksler, R. (**2004**). "Phonological underspecification and mapping mechanisms in the speech recognition lexicon," Brain Lang. **90**, 401–412.
Wiese, R. (**1996**). *The Phonology of German* (Oxford University Press, Oxford).
Wurzel, W. U. (**1970**). *Studien zur Deutschen Lautstruktur* (Studies on the German sound structure) (Akademie Verlag, Berlin).

Estimating vowel formant discrimination thresholds using a single-interval classification task[a)]

Eric Oglesbee[b)] and Diane Kewley-Port
Department of Linguistics and Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 9 September 2008; revised 28 January 2009; accepted 29 January 2009)

Previous research estimating vowel formant discrimination thresholds in words and sentences has often employed a modified two-alternative-forced-choice (2AFC) task with adaptive tracking. Although this approach has produced stable data, the length and number of experimental sessions, as well as the unnaturalness of the task, limit generalizations of results to ordinary speech communication. In this exploratory study, a typical identification task was used to estimate vowel formant discrimination thresholds. Specifically, a signal detection theory approach was used to develop a method to estimate vowel formant discrimination thresholds from a quicker, more natural single-interval classification task. In experiment 1 "classification thresholds" for words in isolation and embedded in sentences were compared to previously collected 2AFC data. Experiment 2 used a within-subjects design to compare thresholds estimated from both classification and 2AFC tasks. Due to instabilities observed in the experiment 1 sentence data, experiment 2 examined only isolated words. Results from these experiments show that for isolated words, thresholds estimated using the classification procedure are comparable to those estimated using the 2AFC task. These results, as well as an analysis of several aspects of the classification procedure, support the viability of this new approach for estimating discrimination thresholds for speech stimuli.
 [DOI: 10.1121/1.3086269]

PACS number(s): 43.71.Es, 43.66.Fe [PEI] Pages: 2323–2335

I. INTRODUCTION

Previous work on examining vowel formant discrimination has often employed a modified two-alternative-forced-choice (2AFC) task with an adaptive tracking procedure (Kewley-Port and Watson, 1994; Kewley-Port *et al.*, 1996; Kewley-Port and Zheng, 1999; Kewley-Port, 2001; Liu and Kewley-Port, 2004b). Establishing thresholds for discriminating formants is important for determining the baseline ability of the auditory system to process small differences between vowels. Results of this research have shown how formant thresholds change when variables such as linguistic context, fundamental frequency, and hearing status were manipulated using a 2AFC task. While this modified 2AFC adaptive tracking task has been successful in estimating stable vowel formant thresholds, repeated stimulus presentations are quite unnatural compared to the single presentation of utterances that occurs in normal communication. Also, the number of trials necessary to obtain stable thresholds is rather large, and consequently listeners have to participate in multiple experimental sessions over the course of several days. Given these drawbacks, the purpose of the current study was to develop a method that used a linguistic labeling task to estimate vowel formant thresholds comparable to those estimated using a modified 2AFC adaptive tracking paradigm.

Although it is a worthy goal to improve experimental procedures for measuring discrimination for any listener, a particular goal in our laboratory has been to understand vowel processing abilities first in young normal hearing (YNH) listeners and then to determine how age, hearing impairment, or first language might degrade those abilities. Because the unnatural 2AFC task might limit generalizing laboratory results to more ordinary communication by elderly hearing-impaired (EHI) or second language (L2) learners of English, we were motivated to develop more natural methods for obtaining thresholds. Reviewing some of our research to date, several studies have used 2AFC procedures to document how the ability to process isolated vowels degrades with typical high-frequency sloping hearing loss. In the first study, Coughlin *et al.* (1998) examined isolated steady-state vowels and found that vowel formant discrimination abilities for low-frequency formants (F1) were the same between EHI listeners and YNH listeners, but that thresholds were significantly elevated for high-frequency formants (F2). In an effort to relate these thresholds to the identification of vowels, moderate correlations were found between the F2 thresholds and vowel identification by EHI listeners, indicating that there is a relationship between reduced ability to correctly identify vowels and poorer discrimination ability.

In two follow-up studies also using isolated steady-state vowels, Richie *et al.* (2003) and Richie and Kewley-Port (2005) reported similar results for young hearing-impaired

[a)]Portions of the data were presented at the 150th Meeting of the Acoustical Society of America (J. Acoust. Soc. Am. **118**, 1929–1930) and the Fourth Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan (J. Acoust. Soc. Am. **120**, 3129).

[b)]Author to whom correspondence should be addressed. Electronic mail: eric.oglesbee@bethelcollege.edu. Present address: Bethel College, Department of Modern Language Studies, Mishawaka, IN 46545.

listeners. Given that such vowel stimuli are far different from those found in fluent sentences, new synthesis techniques using STRAIGHT (Kawahara *et al.*, 1999) were employed by Liu and Kewley-Port (2004b, 2007) to generate nearly natural vowel stimuli for three nine-word sentences. Their study of EHI listeners examined vowel formant discrimination for words in the sentences, as well as computer edited versions that presented only the word or only the isolated vowel from the word. Their results demonstrated a 100% elevation in F2 formant thresholds for EHI compared to YNH listeners for syllables and sentences. Encouraging results using the nearly natural speech stimuli from STRAIGHT led us to another set of experiments examining the relation between vowel thresholds and identification of vowels, this time with L2 learners of American English (AE). Listeners from three L2 languages with different vowel systems heard AE words. Preliminary results reported by Kewley-Port *et al.* (2005) showed little difference between the abilities of L2 and AE listeners to discriminate vowel formants. This is an important outcome, suggesting that more peripheral level processing abilities for vowels are not compromised by one's first language. However, outcomes of these L2 experiments, as well as those with EHI listeners, would be much stronger if the experimental tasks were more ecologically valid. The present study had the ambitious goal of designing more natural experimental procedures that could be used with many listener populations to determine the just noticeable differences in vowel spectra from a simple identification task [i.e., listen to a word (sentence) once and indicate what word was heard]. It should be noted that since the task we have chosen is an identification/classification approach for threshold estimation, it might not be appropriate for populations where large variability has made connecting identification performance to discrimination performance problematic (e.g., cochlear implant users; Iverson, 2003).

The task chosen for this study is a single-interval standard classification task using a continuum of stimuli that vary in one speech parameter, here the frequency of one vowel formant. Classification tasks (also called "labeling" or "identification" tasks) have long been used to examine the relation between category boundaries and discrimination in speech continua [e.g., voice onset time (VOT) continua from /ba/-/pa/, Lisker and Abramson, 1970]. Specifically, the prediction of discrimination performance from a classification task is at the core of research and theory concerning categorical perception (Harnad, 1987). However, there are clear differences between the goals of the current study and those related to categorical perception. For example, categorical perception is concerned with how discrimination performance changes along a classification continuum, especially at the boundary between categories (Repp, 1984), whereas the current study examines discrimination near a category center. Therefore, although the current study examines discrimination using a classification task, our interest is in determining the just noticeable difference between stimuli in contrast to the categorical perception literature that examines the nature of phonetic categories.

Some categorical perception studies have examined how details of the discrimination task alter the observed discrimination function and, therefore, the relationship between discrimination and predictions from the classification task. For example, if the stimulus properties are more steady state (vowel-like) versus rapidly changing (consonant-like), observed discrimination functions are more uniform and performance exceeds predicted discrimination (Pisoni, 1973; Mirman *et al.*, 2004). Also, changes in the discrimination trials from ABX to more sensitive tasks that use less working memory, such as the 4IAX (Pisoni, 1975) or 4I2AFC (Gerrits and Schouten, 2004) tasks, tend to flatten the discrimination function. While some of these issues are certainly related to the present experiments, these tasks all use two or more presentations of the stimuli per trial to measure discrimination. This contrasts with our goal of estimating discrimination thresholds using a single presentation of a stimulus by referencing the underlying psychometric function.

In the current study, principles from "Thurstonian scaling," as implemented in signal detection theory (SDT) (Macmillan and Creelman, 2005, p. 155), are used to find a point on a classification psychometric where thresholds comparable to those found in a 2AFC task can be estimated. Previously established theoretical relationships between performances on a 2AFC adaptive tracking task using a two-down-one-up (2D1U) rule (Levitt, 1971) and a two-category classification task (two-category classification is functionally equivalent to a yes/no decision) are used to provide a range of sensible cumulative d' values for estimating an empirical threshold from the classification psychometric. Thus, two types of experimental tasks were employed in the present study. In experiments 1 and 2 listeners participated in a single-interval classification task. In this task, a stimulus was presented, and listeners categorized the stimulus as one of two response categories (possible response sets were *bid/bed*, *cut/cot*, and *hack/hawk* in the experiments reported here). In experiment 2, the same listeners also used a modified 2AFC adaptive tracking task. This modified 2AFC adaptive tracking task had the same format as the one used in the previous discrimination studies cited earlier, in particular, Kewley-Port and Watson (1994). On each trial listeners heard three stimuli. The first stimulus presented on each trial was always the "standard." The second and third intervals contained the standard and test stimuli played in random order. Consequently, the possible stimulus sequences were ⟨STS⟩ and ⟨SST⟩, where "S" was a fixed standard for each formant continuum and "T" was the test stimulus. Listeners were asked to decide if the second or third interval contained the test stimulus. Using the typical 2D1U adaptive tracking rule, stable performance was obtained after a large number of trials (here 504 per formant) to measure thresholds from mean reversals.

To generate more natural speech stimuli for the classification task, stimuli were synthesized by using STRAIGHT (Kawahara *et al.*, 1999). STRAIGHT is a powerful speech analysis routine implemented in MATLAB that allows experimenters to manipulate amplitude, fundamental frequency, temporal relationships, and spectral characteristics of natural speech tokens. Although many of our earlier vowel threshold

TABLE I. Acoustic characteristics of stimuli used in experiments 1 and 2.

	F1 (Hz)	F2 (Hz)	F3 (Hz)	Midpoint F0 (Hz)	Duration (ms)
bid	345	2196	2961	183	200
bed	485	1927	2918	162	206
cut	668	1561	2692	160	91
cot	829	1432	2616	157	171
hack	818	1906	2530	161	190
hawk	700	1055	2643	162	190

experiments used more controlled vowels (albeit unnatural sounding) created with the KLTSYN synthesizer (Klatt, 1980), Liu and Kewley-Port (2004a) showed that formant thresholds were very similar when estimated using STRAIGHT-synthesized as compared to formant-synthesized speech for analogous vowel stimuli.

In experiment 1, formant discrimination thresholds for vowels in both isolated words and in sentences were estimated using a classification task and then compared to previously collected 2AFC data (Liu and Kewley-Port, 2004b). As a follow-up, experiment 2 implemented a within-subjects design to estimate these thresholds using both a classification and a 2AFC task. Because unstable data were obtained from the sentence stimuli in experiment 1, only isolated words were examined in experiment 2. The isolated word results from the two experiments indicate that there are common processing mechanisms underlying vowel discrimination and classification that allow vowel discrimination performance to be predicted from classification performance. Furthermore, we present an analysis of the relationship between classification function parameters and estimated thresholds for a particular d' criterion, along with some suggested guidelines for implementing the proposed formant discrimination threshold estimation method.

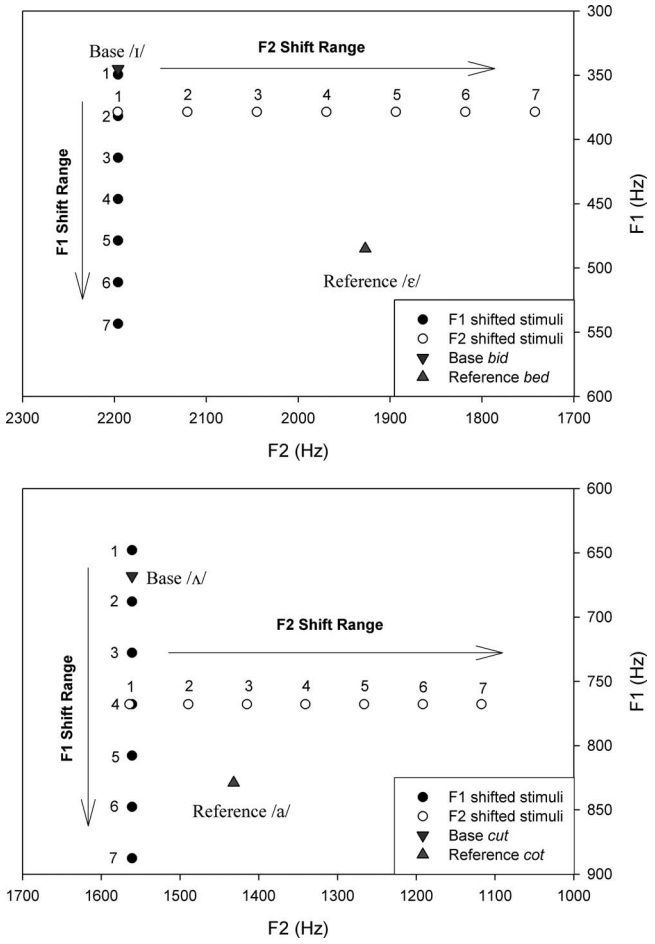


FIG. 1. F1 and F2 values for the four continua used in the forced choice classification (or labeling) tasks in experiments 1 and 2. *Bid/bed* continua are shown in the top panel; *cut/cot* continua are shown in the bottom panel. F1 and F2 shifted continua are indicated by closed and open circles, respectively. Formant values for base stimuli are shown with a downward pointing triangle. Reference stimuli for shifted end points are shown with an upward pointing triangle. Note that for the F2 continua, F1 was first shifted to an intermediate position between the base and reference stimuli.

II. EXPERIMENT 1

A. Method

1. Stimuli

The English vowels /ɪ/, /ɛ/, /ʌ/, and /a/ were chosen for investigation. These vowels were examined in several previous discrimination studies (Kewley-Port and Watson, 1994; Kewley-Port *et al.*, 1996; Kewley-Port and Zheng, 1999; Kewley-Port, 2001; Liu and Kewley-Port, 2004b) that used stimuli generated with either the Klatt (1980) synthesizer or STRAIGHT (Kawahara *et al.*, 1999). In this experiment, using STRAIGHT, four seven-step continua were resynthesized from words taken from recordings of a female speaker who read a list of nine-word sentences that had been designed to have relatively neutral prosody. The speaker for this study was the same one used as a model for synthesis in the studies listed above. The sentences were constructed such that the second, fifth, and eighth word positions within the sentence were filled with combinations of words taken from the minimal pairs *bid/bed*, *cut/cot*, and *hack/hawk*. Listening and acoustic analysis were used to select one prosodically neutral production for each word. Table I contains formant, F0, and vowel duration information for the six selected words. The words *bid* and *cut* were used as base stimuli for resynthesis, while *bed* and *cot* were used as reference stimuli for developing classification continua. *Hack* and *hawk* were used as fillers in the test sentences.

Figure 1 contains a graphical representation of the parameters modified in STRAIGHT across the four classification continua. In each continuum, as in the previous vowel formant discrimination studies cited earlier, a single formant was shifted in equal steps in hertz, while all other acoustic properties were held constant. The direction and magnitude of the formant shifts were determined by informal listening such that the base stimuli *bid* and *cut* would be well-identified as *bed* and *cot*, respectively, for the maximum shift. One strategy was to shift formants to more extreme values than observed in the reference stimuli. Also, to obtain high quality F2 stimuli, F1 was shifted to a somewhat ambiguous value before F2 was shifted. Because the original *bid* token had a relatively high F0, *bid/bed* tokens had F0 lowered by 10% in STRAIGHT's F0 resynthesis routine. To neutralize the duration difference between *cut* and *cot*, duration was lengthened by 20% for the base *cut* stimulus.

The procedure for shifting formant frequencies was a very similar version of the routine employed by Liu and Kewley-Port (2004b). First, a matrix in MATLAB representing the spectrogram (row=frequency; column=time; matrix entry=amplitude) of the base word was obtained by STRAIGHT (Kawahara *et al.*, 1999). Second, to shift a formant peak, the temporal onset and offset of the formant (including transitions) were visually identified. Third, in each time frame (i.e., one spectrum), the bottom of the valleys on either side of the formant peak was selected. Amplitude in the valley opposite of where the formant was shifted was adjusted to be a constant level that was equal to the bottom of the valley, while the shifted peak replaced the remaining spectral values. In most cases, the shift did not result in changes to other formants. A rule was employed where the shifted and original amplitude values were compared, and whichever was larger was retained. This resulted in formants being smoothly merged with one another when shifted to close proximity. Thus, details in the formant peaks were preserved in this procedure, with the valleys only somewhat changed. At last, this modified matrix was reloaded into STRAIGHT, and when necessary, other acoustic modifications such as F0 lowering and temporal expansion were simultaneously implemented [for more details on formant-peak shifting, see Liu and Kewley-Port (2004a) and their Fig. 1].

Sentence stimuli were created by digitally inserting words into the low-predictability sentence frame "Write _____ along with _____ but not _____ please," which was the same frame that was used to elicit the words during recording. Depending on the continuum that was being tested (*bid/bed* or *cut/cot*), the other positions were filled with a word from the pair *hack/hawk* and either *bid/bed* or *cut/cot*. All non-shifted words in the sentence were processed by STRAIGHT with no parameter changes such that the acoustic quality introduced by the STRAIGHT resynthesis was similar for all words tested.

2. Listeners

Indiana University students with normal speech and hearing by self report were paid for their participation in the present study. Of the 35 students tested, data from three listeners (two *bid/bed* F2 and one *cut/cot* F2) were excluded because the listeners had near chance identification of the isolated word end-point stimuli.

3. Procedures

Stimuli were presented to the right ears of listeners, who were seated in a sound-treated IAC booth, via TDH-39 earphones. Stimulus presentation was controlled by TDT System II modules including a 16-bit digital/analog converter, a programmable filter, and a headphone buffer. A low-pass filter with a cutoff frequency of 4500 Hz, and an attenuation level set by the calibration procedure, was loaded into the programmable filter. The standard vowel /ɛ/ was iterated for a duration of 3 s and used for calibration. The sound pressure level (SPL) for the test stimuli measured in a NBS-9A 6-c[3] coupler by a Larson-Davis sound meter (model 2800) using the linear setting was adjusted to be 70 dB SPL, with variation between stimuli from the same continuum less than 1.1 dB SPL.

Listeners were randomly assigned to four groups containing eight listeners each. Each group was exposed to stimuli from only one of the generated continua in both word and sentence conditions (i.e., either F1 shifted *bid/bed*, F1 shifted *cut/cot*, F2 shifted *bid/bed*, or F2 shifted *cut/cot*). Four types of blocks were run: word familiarization, word classification, sentence familiarization, and sentence classification. In each of these tasks, listeners were asked to indicate what word they heard. For isolated words, the possibilities were either *bid/bed* or *cut/cot*, depending on experimental group. For sentences, there were three words identified from the pairs *bid/bed*, *cut/cot*, and *hack/hawk* in each sentence. In the word familiarization blocks, listeners were exposed to the continuum end points in an alternating fashion for 12 trials with feedback for the correct response. In sentence familiarization, listeners heard the nine-word sentence, after which six response buttons appeared containing the words *hack*, *hawk*, *cut*, *cot*, *bid*, and *bed*. The order of presentation of the buttons matched the order of presentation of the word pairs in the sentence. Six sentences with continuum end points were used in the familiarization with feedback provided by the experimenter. The word and sentence classification blocks were similar to their familiarization counterparts, except that all seven stimuli from the continuum were presented multiple times in a random order, without feedback. In the word classification blocks, each stimulus was played nine times, resulting in 63 trials per block. The sentence classification blocks presented each stimulus twice in each sentence position, resulting in 42 trials per block.

Each subject participated in a single experimental session using a fixed protocol of block order. The number of blocks to approximate stable classification performance was determined from pilot experiments. Thus, altogether, four word familiarization, one sentence familiarization, four word classification, and six sentence classification blocks were presented. The session lasted for a total of 75–90 min, depending on a listener's response latency. The last two word classification blocks and last three sentence classification blocks were used for analysis.

B. Results

1. Raw labeling functions

The results for each subject were calculated as the proportion of *bed* or *cot* responses obtained over the 18 presentations of each stimulus. Figures 2 and 3 display a representative set of raw labeling functions for four of the eight conditions. Figure 2 displays data from the *bid/bed* continuum for both F1 and F2 in the word condition. The reasonably consistent performance across listeners exhibited in Fig. 2 are representative of behaviors also observed for *bid/bed* (F1 and F2) in the sentence condition and *cut/cot* (F1 and F2) in the word condition. Listener performance for *cut/cot* (F1 and F2) in the sentence condition was much less consistent than the other conditions and is illustrated in Fig. 3. In general, raw labeling functions were more consistent

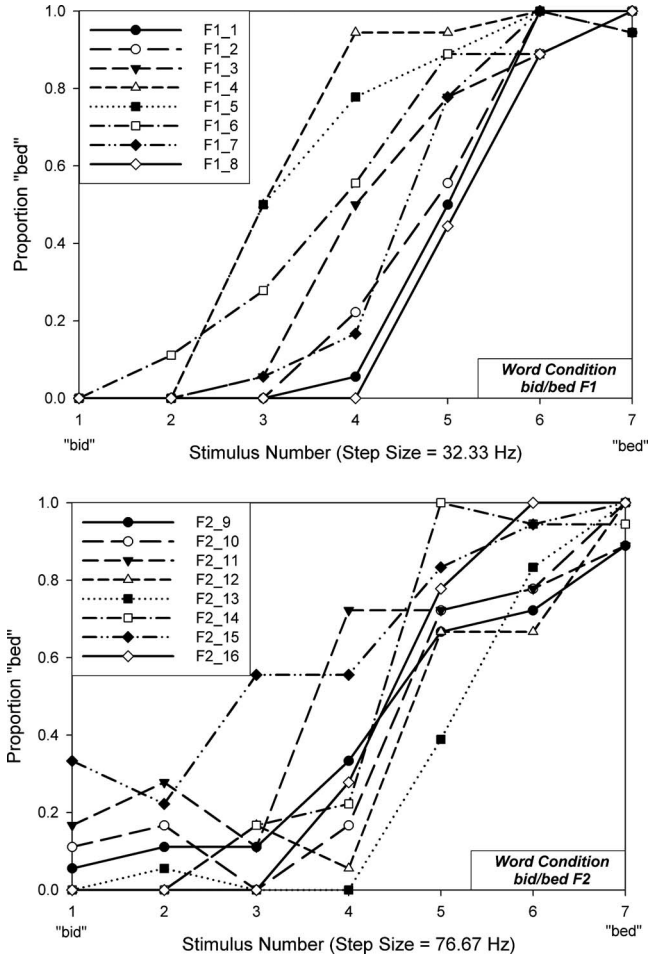


FIG. 2. Raw identification functions for *bid/bed* F1 and F2 stimuli in the word condition (18 judgments per stimulus). Each curve corresponds to a unique listener, numbered 1–16, with an additional identifier (F1 or F2) indicating which experimental group they belonged to. Each continuum contained seven stimuli that were randomized and presented to listeners for identification. The top panel contains results for F1 shifted stimuli; the bottom panel contains results for F2 stimuli. In both panels, stimuli 1 and 7 correspond to the most *bid*- and *bed*-like stimuli, respectively. Stimulus numbers correspond to their respective continua in the top panel of Fig. 1 (i.e., stimulus numbers in the top panel of Fig. 2 correspond to the F1 shift range of the top panel in Fig. 1). Below each panel, the shifted formant peak difference between stimulus steps is given in Hz.

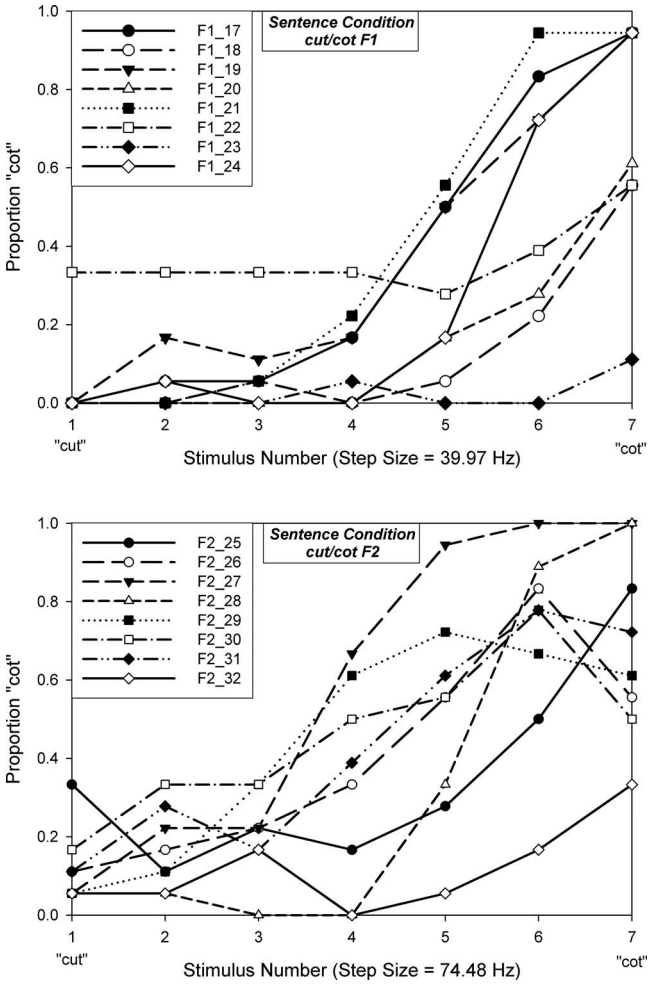


FIG. 3. Raw identification functions for *cut/cot* F1 and F2 stimuli in the sentence condition (18 judgments per stimulus). Each curve corresponds to a unique listener, numbered 17–32, with an additional identifier (F1 or F2) indicating which experimental group they belonged to. Each continuum contained seven stimuli that were randomized and presented to listeners for identification. The top panel contains results for F1 shifted stimuli; the bottom panel contains results for F2 stimuli. In both panels, stimuli 1 and 7 correspond to the most *cut*- and *cot*-like stimuli, respectively. Stimulus numbers correspond to their respective continua in the bottom panel of Fig. 1 (i.e., stimulus numbers in the top panel of Fig. 3 correspond to the F1 shift range of the bottom panel in Fig. 1). Below each panel, the shifted formant peak difference between stimulus steps is given in Hz.

and categorical for the word condition with F1 shifted formants compared to the sentence condition with F2 shifted formants.

2. Threshold calculation method

For discrimination, the concept of a "threshold" is the smallest amount of difference between two sounds that can be reliably detected. In most previous experiments reported by Kewley-Port and her colleagues (Kewley-Port and Watson, 1994; Liu and Kewley-Port, 2004b), vowel formant thresholds were measured as ΔF in hertz using a 2D1U rule in a modified 2AFC adaptive tracking paradigm. These 2AFC thresholds were calculated from average reversals along a stimulus continuum. Mathematically, this corresponds to the 70.7% correct point on a psychometric function plotted between 50% chance and 100% (Levitt, 1971).

In this paper, the term "classification threshold" is used to indicate estimations of discrimination thresholds obtained from the classification task. A difference between thresholds obtained from 2AFC adaptive tracking and classification tasks is that the psychometric function ranges from 50%–100% in a 2AFC task as compared to 0%–100% in a classification task. The goal of the present research was to establish a method to calculate a classification threshold that corresponds well to the 70.7% threshold criterion in a 2AFC discrimination task. This was approached using a SDT implementation of Thurstonian scaling for one-dimensional classification tasks (Macmillan and Creelman, 2005, pp. 114–119). In this approach, a threshold criterion in terms of cumulative d' (measured from a classification continuum end point) is selected. Then, an empirical threshold is calculated by locating the point on the classification psychometric that corresponds to the chosen cumulative d' criterion. In prin-

ciple, the choice of the criterion used to calculate the empirical threshold is arbitrary, although the threshold criteria used in the current experiments were selected with the goal of obtaining results comparable to previously reported vowel formant discrimination thresholds for YNH listeners (Liu and Kewley-Port, 2004b).

The following method was used in the current experiments for deriving an estimation of a discrimination threshold from a two-category classification task (i.e., a classification threshold). It should be noted that this method assumes that the end points of the classification continua were well-identified by a majority of listeners. First, using a least-squares method, a logistic function [see Eq. (1)] was fit to the labeling data,

$$p(s) = 1/(1 + e^{((A-s)/K)}), \quad (1)$$

where A and K are model parameters representing the 50% point and slope of the function, respectively, and s is the stimulus number (in the current experiment s ranged from 1 to 7). Next, the value for $s_{\text{threshold}}$ resulting in $z(p(s)) - z(p(1)) = D$ is determined, where z is the inverse of the normal distribution function and D is the d' value chosen for estimating the empirical threshold (a discussion of the D used in the current experiments appears below). Using the known step size in hertz between stimuli and assuming that the range for s begins at 1, ΔF is calculated using

$$\Delta F = (s_{\text{threshold}} - 1) \times \text{step size} \quad (2)$$

This value, ΔF, is the classification threshold calculated separately for each listener in each condition.

Because the choice of the cumulative d' criterion (i.e., D) associated with a threshold in a one-dimensional classification task is arbitrary, choosing D depends on the goals of the experiment. In the current experiments, the goal was to have the classification thresholds be comparable to those measured in a modified 2AFC adaptive tracking task. Two options presented themselves. One possibility was to collect classification data and then in a *post hoc* analysis determine a value for D which would maximize the correspondence in performance between two tasks. Given the relatively small amount of previous data available, this approach was not pursued. Therefore, we chose a second option, which was to use previously established mathematical relationships between our types of tasks (Macmillan and Creelman, 2005) in an effort to derive a set of candidate values for D and to then determine which of these D-values produced the best overall correspondence between the two tasks. In particular, d' conversions between 2AFC, yes/no, and "reminder" discrimination tasks were selected.

Because a traditional 2AFC adaptive tracking paradigm using a 2D1U rule estimates thresholds at the 70.7% level on the psychometric function, the d' between the 50% chance and 70.7% threshold level was used as the "basic" choice for D [$D_{\text{basic}} = z(0.71) - z(0.5) = 0.553$]. However, SDT also derives a relationship between 2AFC and yes/no tasks (we note that our two-category classification task is functionally equivalent to a yes/no task; i.e., a *bid/bed* decision could be interpreted as either *bid* or *not bid*). In a 2AFC task, d' is $\sqrt{2}$ times that of a yes/no task (Macmillan and Creelman, 2005, pp. 166–168). Thus a second possible cumulative d' criterion is $D_{\text{Y/N}}$, where $D_{\text{Y/N}} = D_{\text{basic}} \times 1/\sqrt{2} = 0.391$. A third possibility also presents itself. Experiment 2 used a modified 2AFC task, with the standard stimulus appearing in the first interval on every trial (⟨STS⟩ and ⟨SST⟩ were the only possible stimulus sequences). One plausible strategy for listeners would be to ignore the third interval and simply attend to the stimuli in the first and second intervals, respectively, since these two intervals contain enough information to make the discrimination decision (⟨ST⟩=test stimulus in the second interval; ⟨SS⟩=test stimulus in the third interval). This strategy is equivalent to interpreting the modified 2AFC task as a "reminder paradigm," where if the listener uses a "differencing" strategy as defined by Macmillan and Creelman (2005), pp. 180–181, the d' for the 2AFC task would be $1/\sqrt{2}$ times that of a reminder task. Thus the third cumulative d' criterion is $D_{\text{remind}} = D_{\text{basic}} \times \sqrt{2} = 0.783$. Analysis of which of these D (D_{basic}, $D_{\text{Y/N}}$, and D_{remind}) result in the best fit to 2AFC thresholds is deferred to experiment 2 where a within-subjects comparison was possible. Given that the outcome of that analysis demonstrated that $D_{\text{remind}} = 0.783$ had the best match, classification thresholds in experiment 1 are reported using this cumulative d' criterion.

It is important to note that the possible relationships between the classification and modified 2AFC tasks were used only as a guide in selecting D. The d' relationship between 2AFC, Y/N, and reminder tasks discussed by Macmillan and Creelman (2005) is in the context of discrimination experiments using identical stimulus sets. More specifically, our experiments differ in two fundamental aspects:

(1) One of our tasks was a two-category *classification* task.
(2) The tasks being compared have different stimulus sets.

Because of these two major differences, it is not clear that previously established relationships between d' in these tasks should hold in our application. We argue, however, that because the current study is exploratory in nature, the above conversion factors suggest a reasonable range for possible values of D.

3. Thresholds for words and sentences

Logistic functions were fit to all raw data functions, and ΔF thresholds [see Eqs. (1) and (2)] were calculated for $D = 0.783$ in Excel using the method described above. Calculated threshold averages and standard deviations for each formant in both the word and sentence conditions are given in Table II. Given the small number of listeners, as well as the unequal variance for word and sentence conditions, a nonparametric sign test (two-tailed) was used to examine separately for each formant whether the word and sentence thresholds within subjects were significantly different. Differences were not found for *bid/bed* F1 ($p < 0.29$) or F2 ($p < 0.73$); however, a statistically significant difference was observed for *cut/cot* F1 ($p < 0.01$) and F2 ($p < 0.01$). For both *cut/cot* F1 and F2, every subject had a higher threshold in the sentence task as compared to the word task. As expected, thresholds for words in sentences were equal to or greater than those for words in isolation.

TABLE II. Formant frequencies, thresholds, and standard deviation of ΔF (Hz) estimated using a classification task for four formants in two phonetic contexts: words and sentences.

	/ɪ/-F1		/ʌ/-F1		/ʌ/-F2		/ɪ/-F2	
Formant frequency	349		648		1564		2196	
	M	s.d.	*M*	s.d.	*M*	s.d.	*M*	s.d.
Word	34.8	6.3	45.4	9.5	73.8	25.1	107.7	29.9
Sentence	38.7	20.9	98.8	68.8	193.7	84.0	137.6	71.8

C. Discussion of experiment 1

1. Comparison to previous research using 2AFC adaptive tracking procedures

One goal for experiment 1 was to develop a method for estimating formant thresholds using a classification task, such that the estimated thresholds would be equivalent to those calculated using a 2AFC paradigm. The previous report of Liu and Kewley-Port (2004b) who determined F1 and F2 formant thresholds based on a 2AFC paradigm for the vowels /ɪ/ and /ʌ/ in /bVd/ contexts, both in isolation and in sentences, was influential in the design of the present experiment. Figure 4 contains individual listener thresholds from the syllable condition of that experiment, as well as the results of the word condition from experiment 1 (additional data from experiment 2 will be discussed later). Thresholds are very similar between Liu and Kewley-Port (2004b) and experiment 1, even though in the current experiment a /kVt/ context was used instead of /bVd/ for testing /ʌ/. Given that the number of listeners per formant was relatively small for the two experiments ($n=4$ and $n=8$) and that the number of listeners was quite different between the experiments, a nonparametric two-tailed Mann–Whitney *U*-test was used to test for differences between the 2AFC and classification data. In the word condition, thresholds for /ɪ/-F1 were found to be significantly different ($U=0, p<0.01$): every threshold estimated from the classification task was lower than those observed in the previous 2AFC data. No statistical differences were found for /ʌ/-F1 ($U=11, p=0.460$), /ʌ/-F2 ($U=13, p=0.682$), and /ɪ/-F2 ($U=15, p=0.934$).

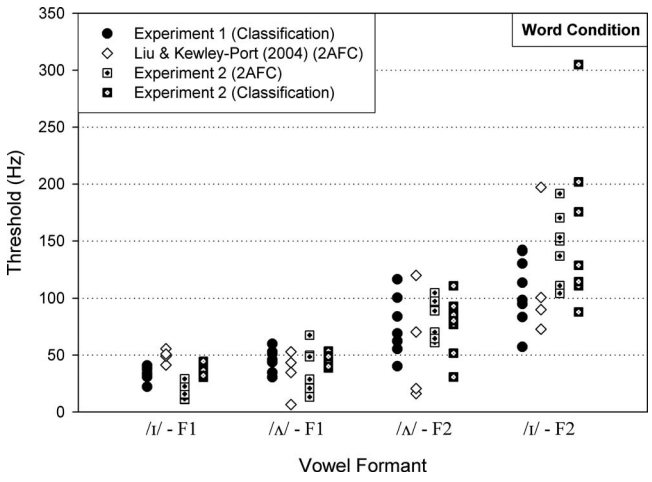


FIG. 4. ΔF for F1 and F2 of the vowels /ɪ/ and /ʌ/ estimated in isolated words. Vowel formants are shown from left to right in order of increasing frequency. Thresholds from individual listeners were estimated using a classification task (experiments 1 and 2) and a 2AFC paradigm (Liu and Kewley-Port, 2004b and experiment 2).

Unfortunately, however, in the sentence condition, thresholds for /ɪ/-F1 and /ʌ/-F1 from the two studies were significantly different ($U=4$, $p<0.05$ for each comparison). While no statistical differences were observed for /ʌ/-F2 ($U=6, p=0.110$) and /ɪ/-F2 ($U=16, p=1$), the range of values obtained in the classification task for /ʌ/-F2 was substantially greater than what was observed in the 2AFC experiment. In general, thresholds obtained from the classification task in the word condition appeared to match the previous 2AFC data better than those estimated from the sentence condition.

2. Positional effects within sentences

Differences in end-point categorization performance and estimated thresholds (Table II) between the word and sentence conditions using *cut/cot* stimuli seemed to indicate that some property of the sentence frame influenced categorization responses and, consequently, estimated thresholds. This contrasts with Liu and Kewley-Port (2004b) where no significant difference in performance was found between the word and sentence conditions. However, in the current study, striking differences for some individual listeners were observed.

Based on the results of Liu and Kewley-Port (2004b) where differences in word position did not significantly affect formant thresholds, by design our identification functions in the sentence condition were originally pooled across the three positions in the sentence frame. However, given the apparent differences between word and sentence conditions in this experiment, the sentence data were further analyzed by sentence position. This analysis using data from the last five of the six sentence blocks revealed one extreme example of position effects (listener F1_22). When *cut/cot* stimuli were placed in the final sentence position, F1_22 exhibited a complete reversal in the identification function. Stimuli that were identified as *cut* in initial and medial sentence positions were labeled as *cot* when presented in the final sentence position, and vice-versa.

Although F1_22 represents the most dramatic case observed among all participants, a visual inspection of the raw position functions across all listeners suggests that stimulus location within sentences had at least a moderate effect on the identification functions of 19 of the other 31 participants. It has been shown that the phonetic content of immediately

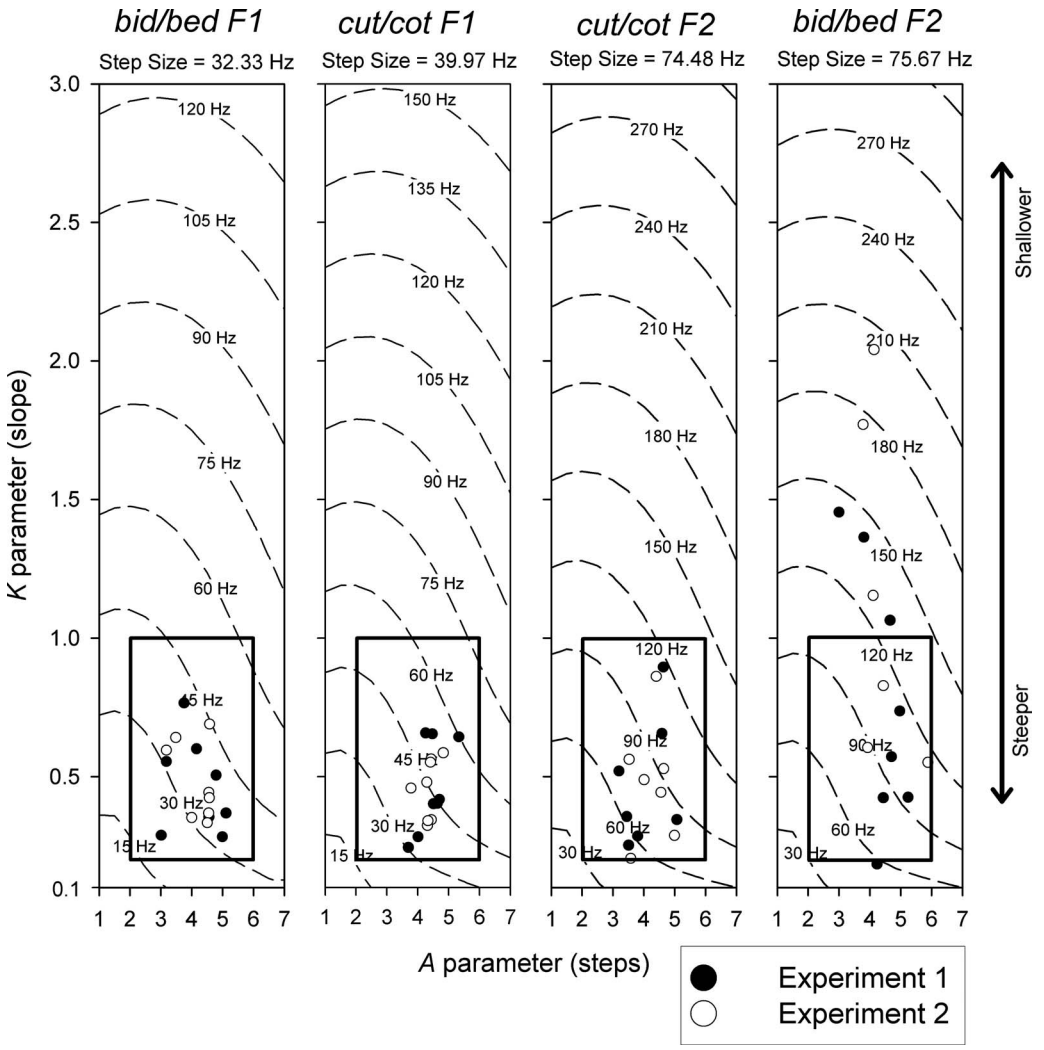


FIG. 5. Contour plots of estimated thresholds as a function of fitted logistic parameters and step size in Hz between stimuli in a classification task using a seven-step continuum. The step at which the fitted logistic reaches the 50% level (A parameter) is plotted on the horizontal axis, and the slope of the fitted logistic (K parameter) is plotted on the vertical axis. Each panel contains contours calculated from (A,K) parameter pairs. Contour lines are labeled with the estimated thresholds based on stimulus step sizes in Hz (shown above each panel) used in experiments 1 and 2. The black boxes in the figure indicates target ranges for (A,K) pairs for obtaining well-behaved threshold functions. Observed (A,K) parameter pairs from experiments 1 (filled circles) and 2 (empty circles) are shown.

adjacent segments can shift vowel perception (Holt *et al.*, 2000). In our results we see a long-distance effect where segments in neighboring words appear to substantially influence categorization. We hypothesize that the likely source of interference is the naturally produced vowel /a/ in the word *not* that appears before the final test position in the frame sentence; however, pursuing this question is beyond the scope of this paper. Consequently, estimating discrimination thresholds using the sentence classification task we designed does not appear to be a viable approach at this time. However, it should be noted that these context effects are not applicable to the isolated word stimuli.

3. Properties of classification threshold estimation procedure

In this section, a preliminary analysis of the general properties of the classification threshold estimation procedure is presented, with design elements and results from the word condition of experiment 1 used to illustrate these general properties. Basic guidelines for designing classification experiments that obtain reliable, stable thresholds are discussed, as well as the generalizability of the procedure. Finally, the success of experiment 1 is evaluated in light of this analysis.

a. General properties Based on the description of the threshold estimation method in Sec. II B 2 of experiment 1, there are three primary factors that affect values calculated for classification thresholds: parameters of the fitted identification function, cumulative d' criterion (D), and step size between stimuli. In the following analysis, the parameters of the logistic function [Eq. (1)] fit in experiment 1 are discussed. These parameters are easily interpretable; namely, A is the 50% point of the function and K is a measure of slope (values close to zero result in steep slopes). It should be noted that the threshold estimation method is general enough that other functions (e.g., Weibull) could be fit to the classification data.

Using the design parameters of experiment 1 (seven-step continua), Fig. 5 contains contour plots of estimated thresh-

olds for $D=D_{remind}=0.783$ as a function of logistic parameters (A parameter on horizontal axis and K parameter on vertical axis) and step size (different panels). Dashed contour lines at 15 and 30 Hz intervals for F1 and F2, respectively, show estimated thresholds. Filled circles mark the thresholds observed in the word condition of experiment 1.

Although these contour plots were created using a specific set experimental parameters, two general observations concerning the threshold estimation method can be made. First, in some regions of the contour plots there is a nonlinear relationship between logistic parameters and estimated thresholds. In particular, when the 50% point of the psychometric occurs early in the function ($A<2$), or when the slope is extremely steep ($K<0.1$) or shallow ($K>2$), the relatively flat portions of the contours indicate that the location of the 50% point has little to no effect on the estimated threshold.

Second, for fixed values of A and K, increasing step size results in specific threshold contours being shifted downward in the $A\times K$ space. For example, the threshold contours in panels 1 (*bid/bed* F1) and 4 (*bid/bed* F2) in Fig. 5 were calculated using step sizes of 32.33 and 75.67 Hz, respectively. Comparing panel 1 to panel 4, the location of the 60 Hz contour line is shifted downward substantially. Note that although this apparently is a strong effect, the experimenter still has a wide range of choices for step size. SDT presumes there is a fixed underlying probability distribution for perceiving stimuli along a continuum. Thus, assuming listener bias remains constant, a reasonable hypothesis is that listener responses will adjust to different step sizes, resulting in sets of logistic parameters that give comparable thresholds using our method. While research is needed in order to verify this hypothesis, the following example demonstrates that the hypothesis is reasonable. Suppose two continua testing the same formant were created from the same initial vowel with step sizes of 32.33 and 39.97 Hz respectively (these are the same step sizes for the *bid/bed* and *cut/cot* F1 panels in Fig. 5). All things being equal, if the same listener were tested on both continua with different step sizes, the slope of the psychometric function for the continuum with a step size of 39.97 Hz should be steeper than the continuum with a step size of 32.33 Hz. Also, because the continua had the same initial vowel, the 50% level of the continuum with the 39.97 Hz steps should be shifted leftward. Thus, the resulting thresholds should be nearly the same even when the physical step size differs.

b. Guidelines for experimental design The threshold contours in Fig. 5 have implications for experimental designs using the new threshold estimation method. Although the contour plots span a large range of A and K parameters, it appears that the range of parameters that should be considered acceptable is much smaller due to the nonlinearities in the threshold contours. In the case of a seven-step stimulus continuum as was used in experiment 1, the black box ($2\leq A\leq 6$; $0.2\leq K\leq 1$) in Fig. 5 represents the set of (A,K) parameter pairs that result in well-behaved threshold contours. This region also corresponds to the set of (A,K) parameter pairs that one would expect from a well-designed two-category identification task. Thus, the black box in Fig. 5 can be used as a *post hoc* evaluation for the appropriateness of classification thresholds estimated from a seven-step continuum.

c. Generalizability Although the above analysis and plots in Fig. 5 illustrate results from a seven-step continuum, preliminary examination of the threshold contour space suggest that the threshold estimation method may be applicable to continua containing 5–20 stimuli. In these cases, the range of A parameters contained within the black box would need to be adjusted; however, the K range should remain fixed. Also, the nature of the classification task suggests that the proposed estimation method may be used for estimating thresholds from any type of stimulus continuum that is designed to span two distinct categories, not just speech. For example, in vision research, thresholds for color categories could be examined. Obviously, further research is needed to examine either the effect of different numbers of steps or the application of the method to other stimuli.

One inherent limitation imposed by using any classification task is that only primary acoustic cues for a contrast can be studied in isolation (e.g., F1 or a complex cue such as VOT). This requirement arises out of the fact that different levels of the cue chosen for study must be capable of causing listener's judgments of category membership to shift reliably. Consequently, this method would not be appropriate for estimating F3 or F4 discrimination thresholds. However, the method would still be applicable if one wanted to pursue "multidimensional discrimination thresholds" where multiple primary and secondary cues are covaried in a single classification continuum, as is typical with complex cues such as VOT. Thus a future study of thresholds for vowels where F1 and F2 covary is in line with our goal of developing a more ecologically valid approach to formant discrimination.

d. Relation to traditional classification measures and predictions of discrimination from classification With respect to classification tasks, the 50% point on the psychometric function is typically the focus of study; however, the proposed threshold estimation method instead focuses on the end points of the psychometric. In order to test whether or not the thresholds estimated in experiment 1 are substantially different from the 50% point, a correlational analysis between estimated thresholds and the location of the 50% point was performed. First, a bark scale transform was applied to the data so that all four formants could be included in the same analysis (Kewley-Port and Zheng, 1998). A very small correlation was found ($r=-0.107$), indicating that classification thresholds are not directly related to the more traditional 50% boundary.

With respect to predicting discrimination performance from a classification task, the slopes of psychometric functions have previously been related to peaks in ABX discrimination functions (Repp, 1984). We note that the current approach incorporates more information in the determination of a discrimination threshold than previous approaches, because it takes into account both the slope of the psychometric *and* the location of the 50% crossover point. Below, in experiment 2, it is demonstrated that these two traditional measures of boundary performance (psychometric slope and 50% crossover point) can be successfully used to estimate within-category (i.e., non-boundary) discrimination thresholds.

TABLE III. Shifted formant step sizes for 2AFC stimuli in experiment 2.

	/ɪ/-F1	/ʌ/-F1	/ʌ/-F2	/ɪ/-F2
Base formant frequency (Hz)	349	648	1564	2196
Step size (Hz)	2.97	5.51	12.51	18.89
% of base frequency	0.85%	0.85%	0.80%	0.86%

e. Evaluation of experiment 1 In the four panels of Fig. 5 the (A,K) parameter pairs obtained for all listeners from the word condition of experiment 1 are plotted with filled circles. For three of the listener groups (*bid/bed* F1, *cut/cot* F1, and *cut/cot* F2), all of the data points fall within the black box discussed above; however, three of the *bid/bed* F2 listeners fall outside the black box, suggesting that the *bid/bed* F2 stimuli were not as well-designed as the others. Overall though, even with the *bid/bed* F2 difficulties, experiment 1 appears to have been well-designed with respect to the properties of the threshold estimation method.

III. EXPERIMENT 2

Given that classification thresholds for words (but not sentences) appear reliable, the purpose of experiment 2 was to directly compare, using a within-subjects design, formant thresholds estimated from a classification task with those calculated using a modified 2AFC adaptive tracking procedure. The motivation for using a within-subjects design was threefold. First, the within-subjects design permitted an evaluation of the three cumulative d' criteria discussed earlier (D_{basic}, $D_{\text{Y/N}}$, and D_{remind}). Second, lingering questions remained concerning the comparability of thresholds estimated using the two methods given the numerous differences in experimental variables between Liu and Kewley-Port (2004b) and experiment 1 (i.e., number of subjects, base stimuli for resynthesis, etc.). Third, because of the large variability observed in experiment 1, we wanted to examine the degree to which performance on the two tasks corresponded within individual listeners.

A. Method

1. Stimuli

For the classification portion of experiment 2, the continua developed in experiment 1 were used, allowing for replication of the results from experiment 1. The continua used in experiment 1 needed to be changed for use in a 2AFC adaptive tracking task because the step size between stimuli was too large. Therefore, four modified formant shifted continua were created for the 2AFC task using the same formant shifting methods employed in experiment 1. Similar to Liu and Kewley-Port (2004b), each continuum contained 15 stimuli (including the standard), and step sizes between stimuli were kept between 0.7% and 1% of the formant frequency of the standard stimulus. For each continuum, the standard was identical to the initial point of the corresponding continuum in the classification task. The specific step sizes for each 2AFC continuum were chosen such that some of the points from the classification continuum would be contained in the 2AFC continuum. Table III contains the formant frequencies for each standard, as well as the step sizes for each continuum.

2. Listeners

Eleven Indiana University students with normal speech and hearing by self report were paid for their participation in the present study. Three listeners did not finish the study due to scheduling difficulties and personal choice. Consequently, their data were excluded.

3. Procedure

Each listener participated in five experimental sessions that were completed within three weeks of the initial session. The first two sessions contained classification blocks, and the last three sessions contained 2AFC blocks. Calibration procedures were the same for experiment 2 as experiment 1.

a. Classification task In experiment 1, listeners were exposed to only one of the four stimulus continua. However, in experiment 2, each listener made classification judgments on all four word continua. For each continuum, two word familiarization and four word classification blocks described in experiment 1 were presented in a fixed order. Two continua were assigned in a quasi-random way for each experimental session. Two blocks with the same target words could not occur on the same test day (i.e., *bid/bed* F1 and F2 could not be presented on the same day). Also, each day had to contain one set of F1 blocks and one set of F2 blocks. With these restrictions, a unique continuum ordering was possible for each of the eight listeners.

Each experimental session during the classification portion of the experiment lasted for 0.5–0.75 h (depending on a listener's response latency). Data for estimating thresholds were typically taken from the last two blocks; however, in cases where one of the last two blocks was substantially different than the other, an earlier similar block was used for analysis.

b. 2AFC task The trials for the four vowel formants (F1 and F2 for both /ɪ/ and /ʌ/) were randomly presented within each block (i.e., under medium stimulus uncertainty). The modified 2AFC paradigm of Kewley-Port and Watson (1994) that was described in Sec. I was employed. Each experimental session lasted for 0.75–1 h with 96 trials in each block. Seven blocks per session were administered for three experimental sessions ($N=21$ blocks). ΔF for each listener was averaged from the mean reversals over the last four blocks in which performance was judged as stable by visual inspection.

B. Results

1. Effect of different cumulative d′ criteria

Figure 6 provides an illustration of the effect of the three cumulative d' criteria ($D_{\text{basic}}=0.553$, $D_{\text{Y/N}}=0.391$, and $D_{\text{remind}}=0.783$) discussed earlier on average classification thresholds. In order to quantify the best match for these three values of D, the 2AFC results were used as the expected value, and sums of squares of the differences were calculated across the four formants. For the F1 stimuli, D_{basic} was the best fit, while D_{remind} was the closest match for the F2 stimuli. Summing the squared differences across all four formants, D_{remind} provided the best overall match and therefore

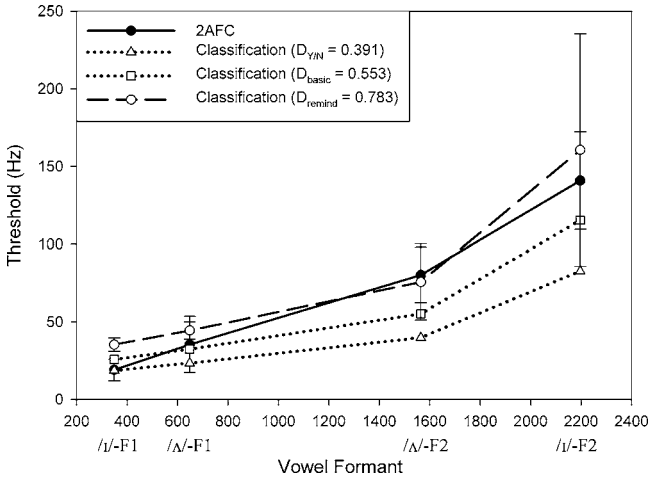


FIG. 6. Average thresholds for /ɪ/-F1 and /ɪ/-F2, and /ʌ/-F1 and /ʌ/-F2, estimated using classification and 2AFC tasks (formant frequency in Hz shown on horizontal axis). Error bars indicating one standard deviation are shown for the 2AFC thresholds, as well as the thresholds estimated using a d' criterion of 0.783 in the classification task. Mean thresholds for d' criteria of 0.391 and 0.553 are shown without error bars.

was used for reporting classification thresholds in experiments 1 and 2. Although this analysis is limited in scope, note that the values of D selected for the current study appear to bracket a reasonable range of values.

2. Classification and 2AFC tasks

In the classification task, raw labeling functions were similar to those observed in experiment 1. Individual thresholds for ΔF were estimated from logistic functions, as described in experiment 1, and are displayed in Fig. 4 (squares) and Table IV.

The average thresholds for the classification and 2AFC tasks were remarkably similar (Fig. 6 and Table IV). Using a repeated measures analysis of variance, a significant main effect for formant was found $[F(3,18)=55.232, p<0.001]$, but not for task $[F(1,6)=0.838, p=0.395]$ or for the task $\times$formant interaction $[F(3,18)=0.290, p=0.832]$. This significant effect of formant is consistent with previous observations that ΔF increases as formant frequency increases (Kewley-Port and Watson, 1994). The non-significant effect of task indicated that the null hypothesis of no difference between the tasks could not be rejected.

3. Threshold estimates within individual listeners

The similarity of the group results does not indicate the relation between classification and 2AFC thresholds within individual listeners. To determine whether the quicker, more natural classification task taps into the same underlying discrimination process as a 2AFC task, several analyses were performed.

First, a Spearman rank order correlation coefficient was used to determine if performance for listeners was ordered from best to worst similarly in the two tasks. Data for /ʌ/-F1 and /ɪ/-F1 conditions were excluded because variability in thresholds for the classification task appeared too small (s.d. $\leq$ 5.5 Hz, Table IV, Fig. 4) for correlations to be meaningful. Correlation coefficients for both /ʌ/-F2 ($N=8, \rho=-0.0238$) and /ɪ/-F2 ($N=7, \rho=0.2143$) indicate that the order of listener performance was not similar across the two tasks. The second way of comparing performance on the two tasks used a two-tailed correlated t-test to evaluate the similarities in individual performance in the F2 conditions. The differences in threshold estimates were not significant for either /ʌ/-F2 ($p=0.717$) or /ɪ/-F2 ($p=0.621$). Taken together, thresholds for individual listeners are generally similar between the two tasks but are not close enough to preserve order from best to worst performance.

Additional analyses of data from all four formants were made after a bark transform removed the effect of the increase in formant frequency on thresholds (Kewley-Port and Zheng, 1998). The bark transformed signed difference between 2AFC and classification thresholds was chosen as the dependent variable for the analyses in order to preserve whether classification thresholds either under- or overestimated the standard 2AFC thresholds, where zero indicates no difference. The average difference between the tasks was −0.063 barks (s.d.=0.163), indicating that individual thresholds were on average slightly lower for the 2AFC than the classification task as hypothesized. While −0.063 barks is small in relation to the 0.10 bark threshold for vowels tested under optimal listening conditions (Kewley-Port and Zheng, 1999; Liu and Kewley-Port, 2004b), the standard deviation of 0.163 is somewhat high.

Overall, there are several ways in which the 2AFC and classification thresholds were comparable. The group thresholds were very similar, with a tendency for classification thresholds to be higher. Although some listeners performed similarly on the two tasks, generally, individual listener thresholds were not ordered similarly from best to worst in the two tasks. Apparently listeners in the two procedures were influenced by separate factors that can result in large differences in individual performance. Thus while classification thresholds cannot be considered a substitute for 2AFC thresholds, it is reasonable to suggest that they would be useful as a performance measure in studies of individual dif-

TABLE IV. Means and standard deviations for classification and 2AFC thresholds estimated in experiment 2.

	/ɪ/-F1		/ʌ/-F1		/ʌ/-F2		/ɪ/-F2	
Formant frequency	349		648		1564		2196	
	M	s.d.	*M*	s.d.	*M*	s.d.	*M*	s.d.
Classification	35.4	4.3	44.9	5.5	75.7	24.6	160.6	74.9
2AFC	19.1	6.8	35.5	18.1	80.2	17.9	140.9	31.4

ferences, because for a given individual, classification thresholds were shown to be reliable in their own right (see next section).

C. Discussion of experiment 2

1. Reliability of word classification thresholds in experiments 1 and 2

Although some of the details of the experiment 2 protocol were different than experiment 1, the word stimuli that were used for the classification portion were identical to those used in the word condition of experiment 1. Therefore, the reliability of thresholds shown in Fig. 4 can be examined between listener groups. Using two-tailed *t*-tests assuming equal variance ($N=8$), no statistically significant differences were found between experiments 1 and 2 for any of the formants: *bid/bed* F1 ($p=0.811$), *cut/cot* F1 ($p=0.813$), *cut/cot* F2 ($p=0.877$), and *bid/bed* F2 ($p=0.088$). We note that variance was high and thresholds were not as stable for /ɪ/-F2. Overall, classification thresholds for words appear reliable across the two different experimental groups.

2. Comparison of 2AFC thresholds with previous research

Given the difference in number of listeners per vowel formant between experiment 2 ($N=8$) and Liu and Kewley-Port (2004b) ($N=4$), a non-parametric two-tailed Mann–Whitney *U*-test was used to compare 2AFC thresholds obtained across experiments. No significant difference was found for /ʌ/-F1 ($U=15, p=0.934$), /ʌ/-F2 ($U=12, p=0.570$), and /ɪ/-F2 ($U=8, p=0.214$); however, a significant difference was found for /ɪ/-F1 ($U=0, p=0.004$). In the case of /ɪ/-F1, the thresholds estimated in experiment 2 ($M=19.1$ Hz, s.d.$=6.8$) were less than half the size of those observed in Liu and Kewley-Port (2004b) ($M=49.2$ Hz, s.d.$=5.9$) (see Fig. 4). A likely source of the threshold difference was that even though the same speaker was used for stimulus resynthesis, different recordings were used as base stimuli in the two experiments. Acoustic analysis showed that the base tokens for /ɪ/ were comparable between the two experiments in all ways except one: the difference in intensity between F1 and F2 (18.2 dB) for the experiment 2 /ɪ/ base stimulus was greater than the difference between F1 and F2 (13.9 dB) in the /ɪ/ base stimulus used in Liu and Kewley-Port (2004b). It is hypothesized that the greater prominence of F1 relative to F2 in experiment 2 made F1 of /ɪ/ in the current experiment more salient, and therefore thresholds were lower. Also, the lower relative intensity of F2 may have contributed to the difficulty of developing well-identified /ɪ/-F2 classification stimuli.

IV. SUMMARY AND CONCLUSIONS

A novel method for estimating discrimination thresholds from a classification task that are comparable to those obtained using a modified 2AFC adaptive tracking task was proposed using principles of Thurstonian scaling as implemented in SDT. In experiment 1, this method was tested using words both in isolation and in sentence context. F1 and F2 thresholds for two vowel continua (/ɪ/ to /ɛ/ and /ʌ/ to /a/) in a classification task were examined, and results showed the following:

- For isolated words, classification thresholds for three of the four formants tested were statistically similar to those from an earlier 2AFC task (Liu and Kewley-Port, 2004b).
- In the sentence context, only two of the four classification thresholds were statistically similar to thresholds reported from a 2AFC task (Liu and Kewley-Port, 2004b).
- The sentence context exhibited strong effects on vowel classification, especially for the /ʌ/-/a/ continua.

The implication of this final point is that while the goal of developing more natural methods for estimating discrimination performance in isolated words was met, an extension of these methods to sentences was not successful. However, in light of the promising results for classification thresholds for words, an analysis of the properties of the threshold estimation method was presented, and general guidelines for the use of this method were proposed. The simpler, more natural classification task, together with reasonably natural word length stimuli, should provide ecologically valid methods for examining vowel perception in general. It may be particularly useful for use with EHI and L2 listeners, although this should be experimentally established first.

In experiment 2, the goal was to further explore the relationship between thresholds estimated from classification and 2AFC adaptive tracking tasks by using a within-subjects design. Given the strong effect the sentence context had in experiment 1, the scope of experiment 2 was restricted to vowel classification in isolated words. The same stimuli from experiment 1 were used for the classification task in experiment 2, and continua for the 2AFC task had additional stimuli resulting in smaller steps between the stimuli used in experiment 1. Results from experiment 2 showed the following:

- The best match between thresholds estimated using the two tasks was found using a cumulative d' criterion (D) of 0.783, which corresponded to interpreting the modified 2AFC task as a "reminder design" in which listeners employed a differencing strategy.
- Thresholds were reliable across different groups of listeners; i.e., thresholds estimated in experiment 2 were comparable to those observed in experiment 1.
- At the group level, no statistical differences were observed between thresholds estimated using the two tasks.
- At the individual level, thresholds estimated using the two tasks were not significantly different; however, rank order correlation results suggested that individual differences between listeners influenced performance on the two tasks.
- Differences between thresholds reported in the current study and those reported in a previous study are probably related to differences in the acoustic properties of the base stimuli used to generate the continua in each study.

In summary, this study described a quicker, simpler threshold estimation method that appears to provide a reasonable approximation of thresholds estimated using a 2AFC

adaptive tracking task. Given that the proposed method is much more natural for speech stimuli and that the amount of listening time required to obtain stable thresholds is substantially less than a 2AFC task, estimating discrimination thresholds from a classification task appears to be a useful option.

ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health Grant No. DC-02229 to Indiana University. Special thanks are due to Morgan Dotts for assisting with the collection and handling of data in the second experiment.

Coughlin, M., Kewley-Port, D., and Humes, L. (**1998**). "The relation between identification and discrimination of vowels in young and elderly listeners," J. Acoust. Soc. Am. **104**, 3597–3607.

Gerrits, E., and Schouten, M. E. H. (**2004**). "Categorical perception depends on the discrimination task," Percept. Psychophys. **66**, 363–376.

Harnad S., ed. (**1987**). *Categorical Perception: The Groundwork of Cognition* (Cambridge University Press, New York).

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722.

Iverson, P. (**2003**). "Evaluating the function of phonetic perceptual phenomena within speech recognition: An examination of the perception of /d/—/t/ by adult cochlear implant users," J. Acoust. Soc. Am. **113**, 1056–1064.

Kawahara, H., Masuda-Kastuse, I., and Cheveigne, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Commun. **27**, 187–207.

Kewley-Port, D. (**2001**). "Vowel formant discrimination: Effects of stimulus uncertainty, consonantal context and training," J. Acoust. Soc. Am. **110**, 2141–2155.

Kewley-Port, D., Bohn, O-S., and Nishi, K. (**2005**). "The influence of different native language systems of vowel discimination and identification," J. Acoust. Soc. Am. **117**, 2399.

Kewley-Port, D., Li, X., Zheng, Y., and Neel, A. (**1996**). "Fundamental frequency effects on thresholds for vowel formant discrimination," J. Acoust. Soc. Am. **100**, 2462–2470.

Kewley-Port, D., and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.

Kewley-Port, D., and Zheng, Y. (**1998**). "Auditory models of formant frequency discrimination for isolated vowels," J. Acoust. Soc. Am. **103**, 1654–1666.

Kewley-Port, D., and Zheng, Y. (**1999**). "Vowel formant discrimination: Towards more ordinary listening conditions," J. Acoust. Soc. Am. **106**, 2945–2958.

Klatt, D. H. (**1980**). "Software for cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Levitt, H. (**1971**). "Transformed up-down method in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lisker, L., and Abramson, A. (**1970**). "The voicing dimension: Some experiments in comparative phonetics," in Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967, pp. 563–567.

Liu, C., and Kewley-Port, D. (**2004a**). "STRAIGHT: A new speech resynthesizer for vowel formant discrimination," ARLO **5**, 31–36.

Liu, C., and Kewley-Port, D. (**2004b**). "Vowel formant discrimination for high-fidelity speech," J. Acoust. Soc. Am. **116**, 1224–1233.

Liu, C., and Kewley-Port, D. (**2007**). "Factors affecting vowel formant discrimination by hearing-impaired listeners," J. Acoust. Soc. Am. **122**, 2855–2864.

Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates, NJ).

Mirman, D., Holt, L. L., and McClelland, J. L. (**2004**). "Categorization and discrimination of non-speech sounds: Differences between steady-state and rapidly-changing acoustic cues," J. Acoust. Soc. Am. **116**, 1198–1207.

Pisoni, D. B. (**1973**). "Auditory and phonetic memory codes in the discrimination of consonants and vowels," Percept. Psychophys. **13**, 253–260.

Pisoni, D. B. (**1975**). "Auditory short-term memory and vowel perception," Mem. Cognit. **3**, 7–18.

Repp, B. H. (**1984**). "Categorical perception: Issues, methods, findings," in *Speech and Language: Advances in Basic Research and Practice*, edited by N. J. Lass (Academic, New York), Vol. **10**, pp. 243–335.

Richie, C., and Kewley-Port, D. (**2005**). "Vowel perception by noise masked normal-hearing young adults," J. Acoust. Soc. Am. **118**, 1101–1110.

Richie, C., Kewley-Port, D., and Coughlin, M. (**2003**). "Discrimination and identification of vowels by young, hearing-impaired adults," J. Acoust. Soc. Am. **114**, 2923–2933.

Speech intelligibility in background noise with ideal binary time-frequency masking

DeLiang Wang[a)]
Department of Computer Science & Engineering and Center for Cognitive Science, The Ohio State University, Columbus, Ohio 43210

Ulrik Kjems, Michael S. Pedersen, and Jesper B. Boldt
Oticon A/S, Kongebakken 9, DK-2765 Smørum, Denmark

Thomas Lunner
Oticon Research Centre Eriksholm, Kongevejen 243, DK-3070 Snekkersten, Denmark and Department of Clinical and Experimental Medicine and Technical Audiology, Linköping University, S-58183 Linköping, Sweden

(Received 4 March 2008; revised 21 January 2009; accepted 27 January 2009)

Ideal binary time-frequency masking is a signal separation technique that retains mixture energy in time-frequency units where local signal-to-noise ratio exceeds a certain threshold and rejects mixture energy in other time-frequency units. Two experiments were designed to assess the effects of ideal binary masking on speech intelligibility of both normal-hearing (NH) and hearing-impaired (HI) listeners in different kinds of background interference. The results from Experiment 1 demonstrate that ideal binary masking leads to substantial reductions in speech-reception threshold for both NH and HI listeners, and the reduction is greater in a cafeteria background than in a speech-shaped noise. Furthermore, listeners with hearing loss benefit more than listeners with normal hearing, particularly for cafeteria noise, and ideal masking nearly equalizes the speech intelligibility performances of NH and HI listeners in noisy backgrounds. The results from Experiment 2 suggest that ideal binary masking in the low-frequency range yields larger intelligibility improvements than in the high-frequency range, especially for listeners with hearing loss. The findings from the two experiments have major implications for understanding speech perception in noise, computational auditory scene analysis, speech enhancement, and hearing aid design. [DOI: 10.1121/1.3083233]

PACS number(s): 43.71.Gv, 43.66.Dc [KWG] Pages: 2336–2347

I. INTRODUCTION

Human speech communication typically takes place in complex acoustic backgrounds with environmental sound sources, competing voices, and ambient noise. It is remarkable that human speech understanding remains robust in the presence of such interference. This perceptual ability is thought to involve the process of auditory scene analysis (Bregman, 1990), by which the auditory system first analyzes a noisy input into a collection of sensory elements in time and frequency, also known as segments (Wang and Brown, 2006), and then selectively groups segments into auditory streams which correspond to sound sources.

It is well known that listeners with hearing loss have greater difficulty in speech perception in background noise. A standard way to quantify speech intelligibility in noise is a speech-reception threshold (SRT), which is the mixture signal to noise ratio (SNR) required to achieve a certain intelligibility score, typically 50%. Hearing-impaired (HI) listeners need 3–6 dB higher SNR than normal-hearing (NH) listeners in order to perform at the same level in typical noisy backgrounds (Plomp, 1994; Alcantara *et al.*, 2003). For speech-shaped noise (SSN) which is a steady noise with a long-term spectrum matching that of natural speech, the SRT increase for HI listeners is from 2.5 to 7 dB (Plomp, 1994). For fluctuating noise or competing speech, the increase is considerably higher (Festen and Plomp, 1990; Hygge *et al.*, 1992; Eisenberg *et al.*, 1995; Peters *et al.*, 1998); for a single competing talker, the increase is as much as 10–15 dB (Carhart and Tillman, 1970; Festen and Plomp, 1990; Peters *et al.*, 1998). Note that, for typical speech materials, a 1 dB increase in SRT leads to a 7%–19% reduction in the percent correct score, and a 2–3 dB elevation creates a significant handicap for understanding speech in noisy listening conditions (Moore, 2007).

Although modern hearing aids improve the audibility and comfort of noisy speech, their ability to improve the intelligibility of noisy speech is unfortunately very limited (Dillon, 2001; Alcantara *et al.*, 2003). Extensive research has been made to develop noise reduction algorithms in order to close the SRT gap between HI and NH listeners. Monaural speech enhancement algorithms, such as Wiener filtering and spectral subtraction, perform statistical analysis of speech and noise and then estimate clean speech from noisy speech (Lim, 1983; Benesty *et al.*, 2005). Although these algorithms produce SNR improvements, they have not led to increased speech intelligibility for human subjects (Levitt, 2001;

[a)]Author to whom correspondence should be addressed. Electronic mail: dwang@cse.ohio-state.edu

 0001-4966/2009/125(4)/2336/12/$25.00

Moore, 2003b; Edwards, 2004). Attempts have also been made to directly enhance speech cues, especially formants which are spectral peaks of speech (Bunnell, 1990; Simpson *et al.*, 1990). This processing results in clearer formant structure; however, listening tests with both NH and HI listeners show little improvement in speech intelligibility (Baer *et al.*, 1993; Alcantara *et al.*, 1994; Dillon, 2001). Unlike monaural speech enhancement, beamforming (spatial filtering) with a microphone array has been demonstrated to achieve significant speech intelligibility improvements, particularly with large arrays (Kates and Weiss, 1996; Levitt, 2001; Schum, 2003). On the other hand, practical considerations of hearing aid design often limit the size of an array to two microphones, and the effectiveness of beamforming degrades in the presence of room reverberation (Greenberg and Zurek, 1992; Levitt, 2001; Ricketts and Hornsby, 2003). Additionally, to benefit from spatial filtering target speech and interfering sounds must originate from different directions.

Recent research in computational auditory scene analysis (CASA) has led to the notion of an ideal binary time-frequency mask as a performance upper bound to measure how well CASA algorithms perform (Wang and Brown, 2006). With a two-dimensional time-frequency (*T-F*) representation or decomposition of the mixture of target and interference, where elements in the representation are called *T-F* units, an ideal binary mask (IBM) is defined as a binary matrix within which 1 denotes that the target energy in the corresponding *T-F* unit exceeds the interference energy by a predefined threshold and 0 denotes otherwise. The threshold is called the local SNR criterion (*LC*), measured in decibels. More specifically, IBM is defined as

$$\mathrm{IBM}(t,f) = \begin{cases} 1 & \text{if } s(t,f) - n(t,f) > \mathrm{LC} \\ 0 & \text{otherwise,} \end{cases}$$

where $s(t,f)$ denotes the target energy within the unit of time t and frequency f and $n(t,f)$ the noise energy in the *T-F* unit, with both $s(t,f)$ and $n(t,f)$ measured in decibels. The mask is considered ideal because its construction requires access to the target and masker signals prior to mixing, and under certain conditions the IBM with $LC=0$ dB has the optimal SNR gain among all the binary masks (Wang, 2005; Li and Wang, 2009). As a separation technique, applying the IBM with $LC=0$ dB to the mixture input retains the *T-F* regions of the mixture where target energy is stronger than interference energy while removing the *T-F* regions where target energy is weaker than interference energy.

Varying *LC* results in different IBMs. Recently, Brungart *et al.* (2006) tested the effects of IBM with different *LC* values on speech mixtures with one target utterance and 1–3 competing utterances of the same talker, where the sound levels of all the utterances are set to be equal. Their experimental results show that, when $0\ \mathrm{dB} \geq LC \geq -12$ dB, IBM produces nearly perfect intelligibility scores, which are dramatically higher than in a control condition where speech mixtures are presented to listeners without processing. They suggest that the choice of $LC=-6$ dB, which lies near the center of the performance plateau, may be better than the commonly used 0 dB *LC* for intelligibility improvement. Furthermore, they attribute the intelligibility improvement to the removal of informational masking which occurs when the listener is unable to successfully extract or segregate acoustically detectable target information from the mixture. Anzalone *et al.* (2006) investigated the intelligibility improvements of a related version of IBM, defined by a comparison between target energy and a threshold rather than a comparison between target energy and interference energy. Using mixtures of speech and SSN, they found that IBM leads to substantial SRT reductions: more than 7 dB for NH listeners and more than 9 dB for HI listeners. In addition they reported that, while NH listeners benefit from ideal masking in both the low-frequency (LF) and high-frequency (HF) ranges, HI listeners benefit from ideal masking only at LFs (up to 1.5 kHz). Li and Loizou (2007) used the IBM to generate "glimpses," or *T-F* regions with stronger target energy, to study several factors that influence glimpsing of speech mixed with babble noise. Their results show that it is important to generate glimpses in the LF to mid-frequency range (up to 3 kHz) that includes the first and the second formant of speech, but not necessary to glimpse a whole utterance; high intelligibility is achieved when the listener can obtain glimpses in a majority of time frames. More recently, Li and Loizou (2008b) extended the findings of Brungart *et al.* (2006) to different types of background interference, including speech babble and modulated SSN. Moreover, they evaluated the impact of deviations from the IBM on intelligibility performance and found that there is a gradual drop as the amount of mask errors increases. A subsequent study by Li and Loizou (2008a) shows that NH listeners obtain significant intelligibility improvements from IBM processing with as few as 12 frequency channels, and IBM processing in the LF to mid-frequency range that includes the first and the second formant appears sufficient.

In this paper, we evaluate the effects of IBM processing on speech intelligibility with two kinds of background noise: SSN and cafeteria noise, using both NH and HI listeners. While SSN is commonly used in the literature, the cafeteria noise we use contains a conversation between two speakers in a cafeteria background and it resembles the kind of noise typically encountered in everyday life. Our study adopts the standard IBM definition with a comparison between target and interference and measures speech intelligibility by SRT at the 50% level. As suggested by the findings of Brungart *et al.* (2006), we set *LC* to −6 dB in IBM construction. Intrigued by the observation of Anzalone *et al.* (2006) that HI listeners derive little benefit from IBM in the HF range, we conduct an experiment to test whether ideal masking in the HF range is indeed not important for HI subjects. Unlike Anzalone *et al.* (2006) who applied a constant gain to compensate for the hearing loss of their HI subjects, we apply gain prescriptions to fit individual HI listeners.

In what follows, Sec. II details IBM processing. Section III describes an experiment that tests the effects of ideal masking on mixtures of speech with SSN or cafeteria noise. Section IV describes an experiment that compares the effects of ideal masking in LF, HF, and all-frequency (AF) ranges. Further discussion is given in Sec. V. Finally, Sec. VI concludes the paper.

II. IDEAL BINARY MASKING

The concept of IBM in CASA is directly motivated by the auditory masking phenomenon which, roughly speaking, refers to the perceptual effect that a louder sound renders a weaker sound inaudible within a critical band (Moore, 2003a). So keeping noise in *T-F* units with stronger target energy as done in the standard IBM definition with 0 dB *LC* should not reduce speech intelligibility, and this is indeed what was found by Drullman (1995). On the other hand, IBM processing removes all the *T-F* units with stronger interference energy as the target energy in these units is assumed to be masked by the interference. Removing these masker-dominated units also serves to remove informational masking, which is a dominant factor for reduced speech intelligibility in speech and other modulated maskers (Brungart, 2001). Hence IBM processing, as a form of ideal time-frequency segregation, is expected to yield larger speech intelligibility improvement in a modulated noise condition than in a steady noise condition (Brungart *et al.*, 2006).

Like earlier studies (Brungart *et al.*, 2006; Anzalone *et al.*, 2006), we use a gammatone filterbank to process a stimulus and then time windowing to produce a cochleagram which is a two-dimensional *T-F* presentation (Wang and Brown, 2006). Specifically, we use a 64-channel filterbank that is equally spaced on the equivalent rectangular bandwidth (ERB) rate scale with center frequencies distributed from 2 to 33 ERBs (corresponding to 55–7743 Hz). The bandwidth of each filter is 1 ERB. We note that this filterbank is similar to the one used in Anzalone *et al.* (2006) whereas Brungart *et al.* (2006) used a 128-channel filterbank covering the frequency range of 80–5000 Hz. The response of each filter is divided into 20 ms frames with a frame shift of 10 ms, hence generating a two-dimensional matrix of *T-F* units. The cochleagram of a stimulus is simply the two-dimensional graph of response energy within all the *T-F* units. For a given mixture of target signal and background noise, the IBM is calculated by comparing whether the local SNR within a *T-F* unit is greater than *LC*. As mentioned in Sec. I, we fix $LC=-6$ dB in this study as suggested by Brungart *et al.* (2006). Such a choice of negative *LC* retains certain *T-F* units where the target energy is weaker but not much weaker than the interference energy, in accordance with Drullman's observation that weaker speech energy below the noise level still makes some contribution to speech intelligibility (Drullman, 1995). Indeed, a pilot test with 0 dB *LC* indicates that SRT improvements are not as high as those produced with $LC=-6$ dB. More generally, in order to produce large auditory masking, the masker needs to be stronger than the masked signal (Moore, 2003a).

Given an IBM, the waveform output of IBM can be resynthesized from the mixture input by weighting the mixture cochleagram by the IBM and correcting phase shifts introduced during gammatone filtering (see Wang and Brown, 2006). Such an output can then be played to a listener as a stimulus in our experiments. Figure 1 illustrates IBM for a mixture of a speech utterance and a cafeteria background. The SNR of the mixture is 0 dB. Figure 1(a) shows the cochleagram of the target speech, Fig. 1(b) that of the background noise, and Fig. 1(c) that of the mixture. Figure 1(d) displays the IBM with $LC=-6$ dB, and Fig. 1(e) the cochleagram of the resynthesized result of ideal masking with the IBM in Fig. 1(d). The ideally masked mixture in Fig. 1(d) is clearly more similar to the target speech shown in Fig. 1(a) than the original mixture shown in Fig. 1(c) is. As a comparison, Fig. 1(f) shows the IBM with $LC=0$ dB, and Fig. 1(g) the cochleagram of the corresponding ideal masking output. With the increased *LC*, the IBM has fewer 1's and retains less mixture energy.

III. EXPERIMENT 1: EFFECTS OF IDEAL BINARY MASKING ON SPEECH-RECEPTION THRESHOLD

This experiment was designed to quantify the SRT effects of IBM for both NH and HI listeners. Sentences from the Dantale II corpus (Wagener *et al.*, 2003) were used as target speech, and tests were conducted with two different backgrounds: SSN and cafeteria noise.

A. Methods

1. Stimuli

The Dantale II corpus (Wagener *et al.*, 2003) comprises sentences recorded by a female Danish speaker. Each sentence has five words with a fixed grammar (name, verb, numeral, adjective, and object), for example, "Linda bought three lovely presents" (English translation). Each word in a sentence is randomly chosen from ten equally meaningful words. As a result, recognizing a subset of words in a sentence does not help with the recognition of the remaining words. There are a total of 15 test lists, and each list has ten sentences with no repeating word. There are a few seconds of silence between sentences within each list to allow a listener time to report what has been heard. Similar to the Swedish sentence test (Hagerman, 1982), the closed set corpus was designed for repeated use, and training effects are minimal after familiarization with a few lists (Wagener *et al.*, 2003). We use the speech-shaped noise included with the Dantale II corpus, which is produced by superimposing the speech material in the corpus. The cafeteria noise employed is a recorded conversation in Danish between a male and female speaker that took place in a cafeteria background (Vestergaard, 1998). To emphasize temporal modulation effects, the long-term spectrum of this noise was shaped to match that of the Dantale II speech material (Johannesson, 2006). Target speech and background noises are all digitized at 20 kHz sampling frequency.

A speech utterance and a background noise are first processed separately by a 64-channel gammatone filterbank (see Sec. II), which produces a flat frequency response within the frequency range of the filterbank. Filter responses are then windowed into 20 ms rectangular frames with a 50% overlap between consecutive frames, resulting in a two-dimensional cochleagram. This 100 Hz frame rate is frequently used in speech processing (Rabiner and Juang, 1993). For a given mixture of a Dantale II list and a background noise, the mixture SNR is calculated during the intervals that contain speech energy. To account for the forward masking of the continuously present noise that occurs between two consecu-

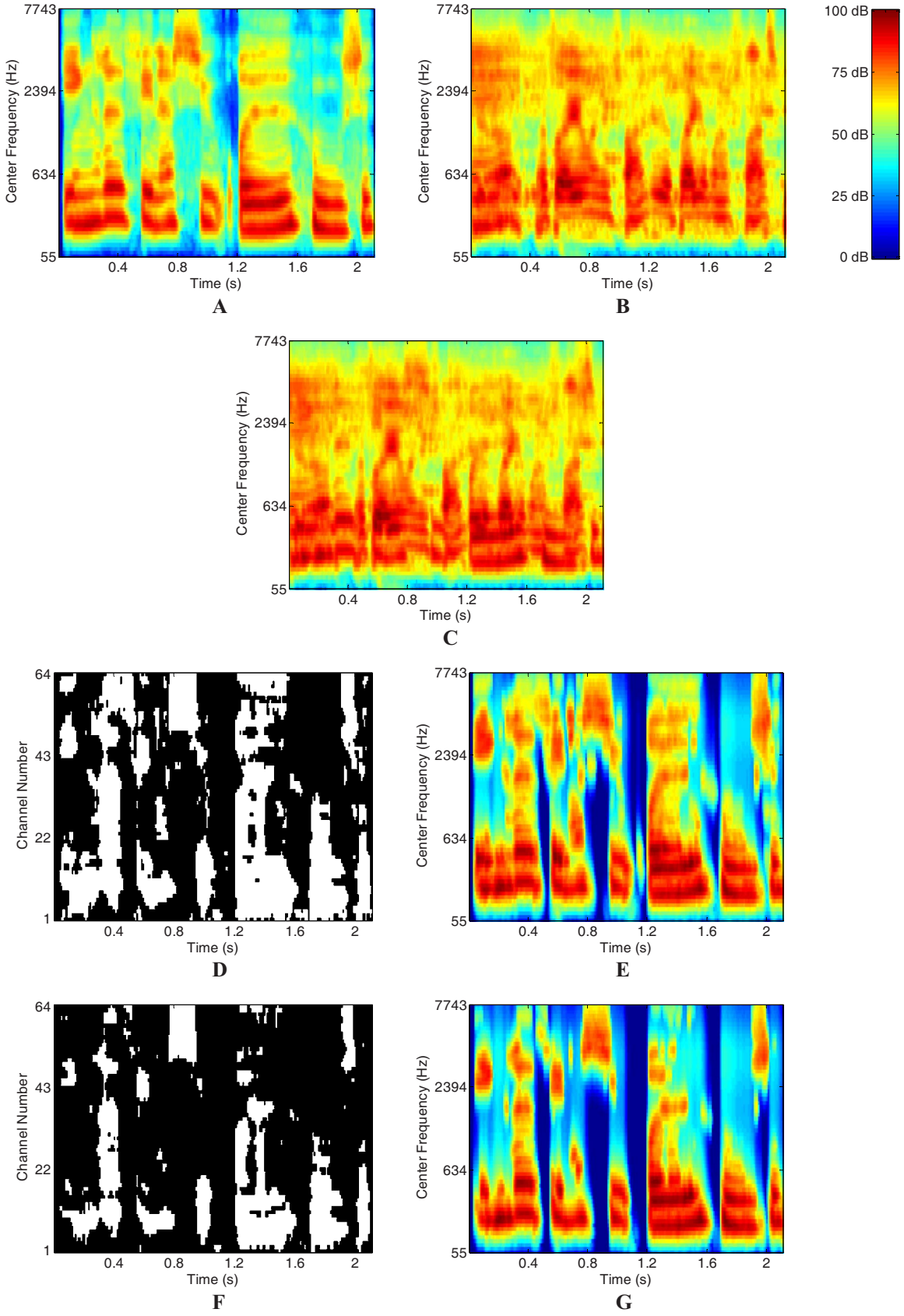


FIG. 1. (Color online) Illustration of IBM (A) Cochleagram of a target speech utterance. (B) Cochleagram of a cafeteria background. (C) Cochleagram of a 0 dB mixture of the speech and the background shown in A and B. (D) IBM with $LC=-6$ dB, where 1 is indicated by white and 0 by black. (E) Cochleagram of the segregated mixture by the IBM in D. (F) IBM with $LC=0$ dB. (G) Cochleagram of the segregated mixture by the IBM in (F).

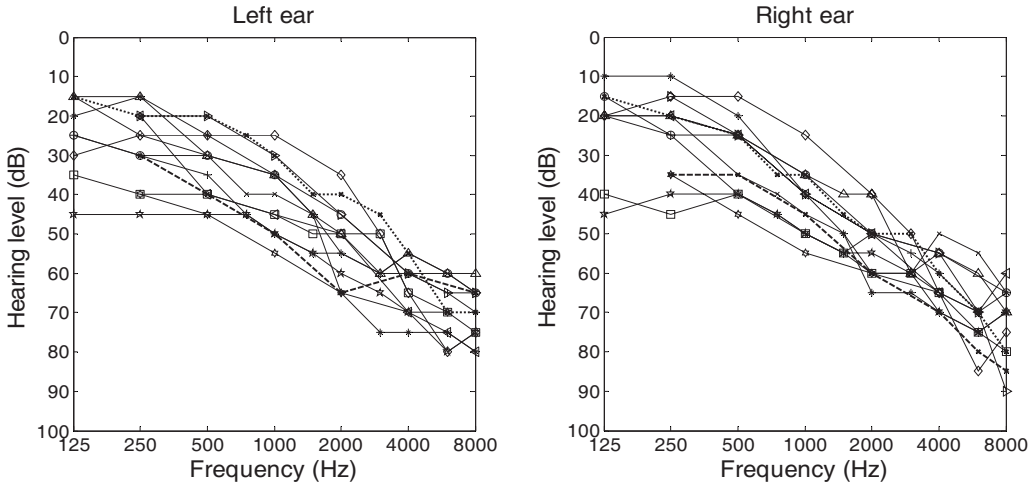


FIG. 2. Audiograms of the 13 HI listeners who participated in the experiments. The dashed line indicates the subject who only participated in Experiment 1, and the dotted line the subject who only participated in Experiment 2.

tive sentences (Moore, 2003a), a period of 100 ms is added before the onset of a sentence for mixture SNR calculation. For a mixture input with a specified SNR, IBM is constructed from the cochleagrams of the target speech and the background noise with *LC* fixed at −6 dB. The IBM is then used to resynthesize a waveform stimulus from the mixture cochleagram. Note that, as a result, the masker signals in between sentences are removed by IBM processing because during such intervals there is only masker energy.

As control conditions, mixtures of speech and background noise were also presented to listeners without segregation. To incorporate filtering effects and any distortions that might be introduced during cochleagram analysis and synthesis, a mixture in an unsegregated condition is processed through an all-1 binary mask or the IBM with the *LC* of negative infinity, therefore including all the *T-F* units in the resynthesis.

2. Listeners

A total of 12 NH listeners and a total of 12 listeners with sensorineural hearing loss participated in this experiment. All subjects were native Danish speakers. The NH listeners had hearing thresholds at or below 20 dB HL from 125 Hz to 8 kHz, and their ages ranged from 26 to 51 with the average age of 37. The NH listeners had little prior experience with auditory experiments, and were not informed of the purpose or design of the experiment.

The 12 HI listeners had a symmetric, mild-to-moderate, sloping hearing loss. The audiograms of these listeners are shown in Fig. 2. They had an age range from 33 to 80 with the average age of 67. All the HI listeners were experienced hearing aid wearers. The tests were performed with their hearing aids taken off, and compensation was applied to each HI subject individually. Specifically, a gain prescription was computed from an individual's audiogram using the NAL-RP procedure (Dillon, 2001), and then used to produce amplification with appropriate frequency-dependent shaping. The hearing losses in the left ear and the right ear were compensated for separately. The subjects had participated in Dantale II listening tasks before, but were not told of the purpose and design of this experiment.

3. Procedure

There are a total of four test conditions in this experiment: two ideal masking conditions with SSN and cafeteria noise and two control conditions with unsegregated mixtures. Three Dantale II lists with a total of 30 sentences were randomly selected from the corpus for each test condition. Subjects were instructed to repeat as many words as they could after listening to each stimulus that corresponded to one sentence, and they were not given any feedback as to whether their responses were correct or not. To familiarize them with the test procedure, subjects were given a training session at the beginning of the experiment by listening to and reporting on three lists of clean sentences. The order of the four conditions was randomized but balanced among the listeners (Beck and Zacharov, 2006). A subject test with the four conditions and a training session together took less than 1 h, and a short break was given roughly halfway through the test.

The Dantale II test employs an adaptive procedure in order to find the 50% SRT. The procedure is to present test sentences at SNR that is continuously adapted according to the number of correctly reported words in the previous sentence (Hansen and Ludvigsen, 2001). In a test condition with 30 sentences, the first 10 sentences are used to reach a steady 50% SRT level and the final SRT is determined by averaging the SNR levels for the last 20 sentences.

Speech and noise were both set to the same initial sound pressure level (SPL) for NH listeners. For HI listeners, the initial SPL of speech was set to 5 dB higher than the noise SPL in Experiment 1, and to the same SPL of noise in Experiment 2. In unsegregated conditions, the noise level was fixed while the speech level was adjusted during the adaptive procedure. In ideal masking conditions, as input SNR drops IBM becomes sparser with fewer 1's. To ensure that ideally masked stimuli remain audible at very low SNRs, the speech

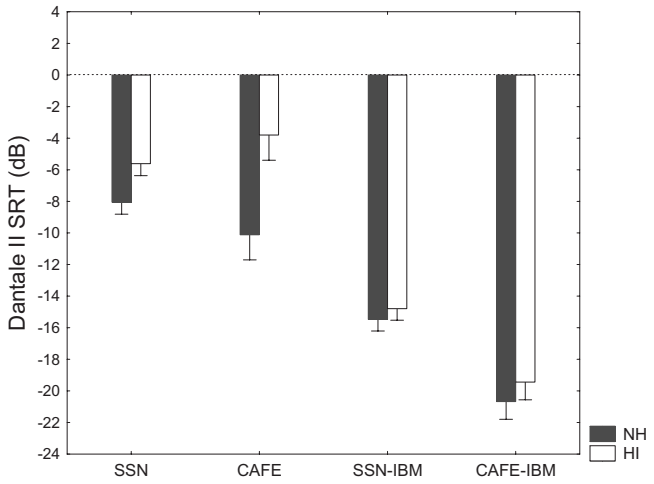


FIG. 3. SRTs for different conditions of Experiment 1 for NH and HI listeners. A more negative SRT corresponds to better performance. Error bars indicate 95% confidence intervals of the means.

level was fixed while the noise level was adjusted in all IBM conditions. As a result, with fewer retained *T-F* units their sound levels became higher even though the levels of the speech signals within these units were unchanged, and the loudness of a processed mixture was thus kept within a small range. This way of adjusting input SNR ensured that the stimuli in all four conditions were comfortably audible.

During a test, a subject was seated in a sound attenuating booth. Test stimuli were generated using the built-in sound card (SoundMAX) in a control computer (IBM ThinkCenter S50) and then presented diotically to a listener through headphones (Sennheiser HD 280 Pro). For HI listeners, an external amplifier (Behring Powerplay HA4000) was used to increase the sound level so that the stimuli within the test range were all audible and yet not uncomfortably loud. The amplification level was adjusted once for each HI listener before the test began.

4. Statistical analysis and power

During the planning phase of the study, the experiment was statistically powered to detect a within-subject between-condition difference of 1.0 dB on mean scores across conditions on the Dantale II test described subsequently for $p<0.05$ at 80% power. This required at least ten complete data sets per condition. Analysis of variance (ANOVA) was performed on all of the data from NH and HI subjects, with within-subject factors of type of processing (IBM or unsegregated) and of type of noise (SSN or cafeteria noise), and a between-subject factor of subject type (NH and HI). *Post hoc* tests were the Bonferroni test and/or the Fisher least-significant-difference (LSD) test, applied where appropriate. The Bonferroni test was used as the most conservative test to indicate differences between means, while the Fisher LSD test was used as the most conservative test for a null result. All statistics were performed using STATISTICA version 7 (StatSoft, 2007).

B. Results and discussion

Figure 3 shows the SRT results of all four test conditions: SSN, CAFE, SSN-IBM, and CAFE-IBM, for both NH and HI listeners. For NH listeners, the mean SRT for unsegregated mixtures with SSN (SSN) is −8.15 dB, for unsegregated mixtures with cafeteria noise (CAFE) is −10.25 dB, for ideal masking with SSN (SSN-IBM) is −15.56 dB, and for ideal masking with cafeteria noise (CAFE-IBM) is −20.70 dB. The ANOVA for NH subjects showed that the main effects of processing type and noise type were significant [$F(1,11)=606.1$, $p<0.001$, and $F(1,11)=78.1$, $p<0.001$, respectively], and there was also a significant interaction between processing type and noise type [$F(1,11)=32.3$, $p<0.001$]. The Bonferroni *post hoc* tests indicated that all means were significantly different [$p<0.005$] from one another. The results show that ideal masking leads to lower (better) SRT compared to unsegregated mixtures regardless of background noise, that the cafeteria background yields a lower SRT than the SSN, and that ideal masking has a greater effect on the cafeteria background. The SRT for the unsegregated SSN condition is comparable to the reference level of −8.43 dB for the Dantale II task (Wagener *et al.*, 2003). The lower SRT for the cafeteria background is consistent with previous studies showing that NH listeners exhibit higher intelligibility in fluctuating backgrounds (Festen and Plomp, 1990; Peters *et al.*, 1998).

For the SSN background, IBM produces an average SRT improvement of 7.4 dB. This level of improvement is consistent with what was found by Anzalone *et al.* (2006) using the HINT test (Nilsson *et al.*, 1994), but higher than the 5 dB improvement reported by Brungart *et al.* (2006) using the CRM task (Bolia *et al.*, 2000). The main difference between our experiment and Brungart *et al.* (2006) lies in different *LC* values: their test uses 0 dB *LC* whereas *LC* is set to −6 dB in our study. As reported in Brungart *et al.* (2006) the choice of $LC=-6$ dB seems better than $LC=0$ dB in terms of speech intelligibility (see also Sec. II).

For the cafeteria background, ideal masking lowers SRT by 10.5 dB on average, which represents a larger gain than for the SSN background. Unlike SSN, the cafeteria background contains significant spectral and temporal modulations which contribute to better intelligibility in the unsegregated condition. We stress that the larger SRT improvement for this background is achieved on top of the better performance of listening to unsegregated mixtures.

For HI listeners, the mean SRTs are −5.61, −3.80, −14.79, and −19.44 dB for the SSN, CAFE, SSN-IBM, and SSN-CAFE conditions, respectively. The ANOVA where both NH and HI subjects were included showed that the main effects of subject type, processing type, and noise type were significant [$F(1,22)=17.2$, $p<0.001$; $F(1,22)=1959.0$, $p<0.001$; and $F(1,22)=100.6$, $p<0.001$, respectively], and there were also significant interaction effects between-subject type and processing type, subject type and noise type, and processing type and noise type [$F(1,22)=49.9$, $p<0.001$; $F(1,22)=19.2$, $p<0.001$; and $F(1,22)=163.9$, $p<0.001$ respectively], as well as a three-way interaction between subject type, processing type, and noise type [$F(1,11)=19.7$, $p<0.001$]. The Bonferroni as well as the Fisher LSD *post hoc* tests on the three-way interaction indicated that all means were significantly different ($p<0.006$)

except for the SSN-IBM and CAFE-IBM conditions where the differences between NH and HI listeners were insignificant ($p > 0.05$). The *post hoc* results show that ideal masking produces lower SRT compared to unsegregated mixtures regardless of noise type, and has a greater effect for the cafeteria background. No difference, however, was revealed between the NH subjects and the HI subjects in the two IBM conditions by either the more conservative Bonferroni test or the less conservative Fisher LSD test. The elevated levels of SRT in the two unsegregated conditions show that HI listeners perform worse in speech recognition in noisy environments, and the levels of SRT increment, 2.5 dB for the SSN condition and 6.5 dB for the CAFE condition, are broadly compatible with previous findings of HI listeners' increased difficulty in speech understanding in noise (see Sec. I). IBM lowers SRT substantially. The SRT gain resulting from ideal masking is 9.2 dB for the SSN background, and this level of improvement is compatible with that reported in Anzalone *et al.* (2006). For the cafeteria background, ideal masking produces a very large SRT improvement of 15.6 dB.

By comparing NH and HI results in Fig. 3, it is clear that HI listeners benefit from ideal masking even more than NH listeners, particularly for the cafeteria background. The results suggest that, after IBM processing, the intelligibility performance is comparable for HI and NH listeners in both SSN and cafeteria backgrounds. It is remarkable that the speech intelligibility of HI listeners becomes statistically indistinguishable from that of NH listeners after ideal masking.

IV. EXPERIMENT 2: EFFECTS OF BAND-LIMITED IDEAL BINARY MASKING ON SPEECH-RECEPTION THRESHOLD

The results of Experiment 1 show large SRT improvements resulting from IBM processing. As mentioned in Sec. I, a main finding reported by Anzalone *et al.* (2006) is that, while NH listeners benefit from IBM in both the LF and HF ranges, HI listeners benefit from ideal masking only in the LF range. This finding is significant because it suggests that, to alleviate the hearing loss of HI listeners, one need not worry about performing *T-F* masking in the HF range; speech segregation at HFs tends to be more difficult than at LFs (Wang and Brown, 2006). Although their interpretation based on the upward spread of masking is reasonable, the fact that they apply constant amplification with no spectral shaping to compensate for the sloping hearing loss of their subjects may suggest a simpler interpretation: the lack of the IBM benefit in the HF range may be partially accounted for by the potentially less compensated hearing loss at HFs. Experiment 2 was primarily designed to assess the importance of IBM processing at HFs for HI listeners as compared to NH listeners. In this experiment, we compensated for the hearing loss of individual listeners based on their audiograms. We compare the intelligibility performance in three setups: IBM in the LF range only, ideal masking in the HF range only, and ideal masking in the AF range. Both SSN and cafeteria backgrounds are used. Consequently, there are a total of six test conditions in this experiment.

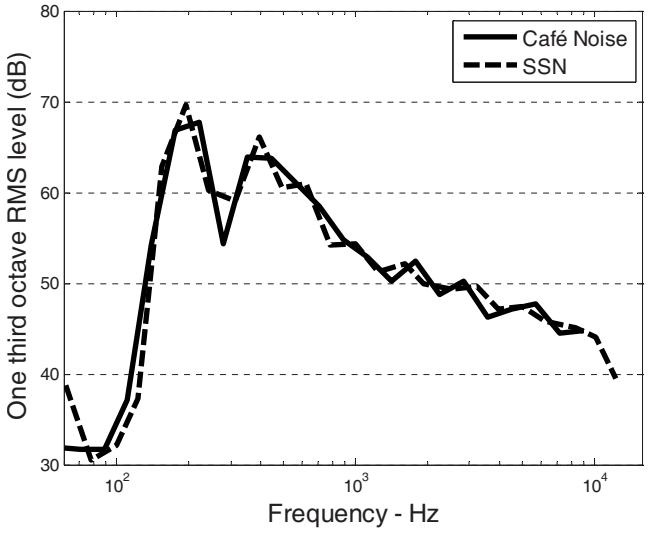


FIG. 4. Long-term spectrum of the SSN in Dantale II (redrawn from Wagener *et al.*, 2003). The spectrum is expressed as root mean square levels in one-third octave bands. Also shown is the long-term spectrum of the cafeteria noise.

A. Methods

1. Stimuli

As in Experiment 1, Dantale II sentences were used as target speech, and SSN and cafeteria noise were used as two different backgrounds. The IBM processing in the AF condition was the same as in Experiment 1. For the LF condition, the same IBM processing as in Experiment 1 was used in the lower 32 frequency channels while an all-1 mask was applied to the higher 32 frequency channels. This way of processing produces no segregation in the HF range. In the HF condition, the reverse was done: IBM was applied to the higher 32 channels while an all-1 mask was applied to the lower 32 channels (hence no segregation in the LF range). This equal division of the 64-channel gammatone filterbank yields a frequency separation boundary approximately at 1.35 kHz. Note that this boundary separating LF and HF ranges is a little lower than the 1.5 kHz boundary used in Anzalone *et al.* (2006). Our choice was partly motivated by the consideration that both the speech material and the SSN in the Dantale II corpus have energy distribution heavily tilted toward LFs so that IBM processing below 1 kHz likely removes significantly more noise than IBM processing above 1 kHz. The long-term spectrum of the SSN (Wagener *et al.*, 2003) is shown in Fig. 4, along with the long-term spectrum of the cafeteria noise. With the 1.5 kHz boundary, the NH results from Anzalone *et al.* (2006) show that the SRT in their LF condition is a little lower than the SRT in their HF condition.

2. Listeners

12 NH listeners and 12 HI listeners participated in this experiment. The pool of NH listeners was the same as that participated in Experiment 1 except for one. This substitution lowered the average age from 37 to 36 without altering the age range. The pool of HI listeners also remained the same as in Experiment 1 except for one. This substitution (see Fig.

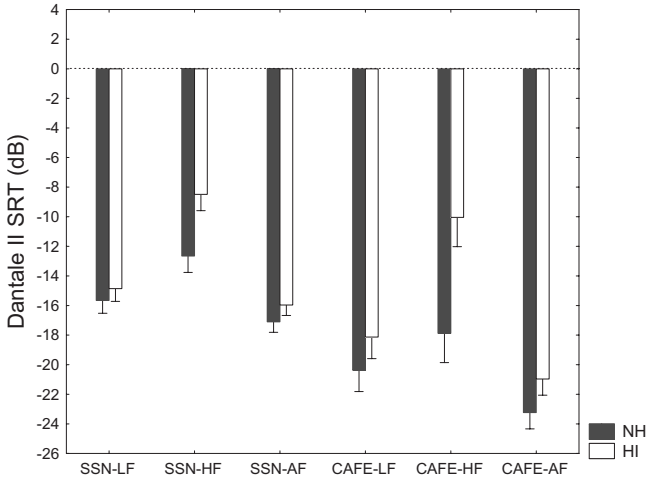


FIG. 5. SRTs for different conditions of Experiment 2 for NH and HI listeners. Error bars indicate 95% confidence intervals of the means.

2), plus a listener whose birthday occurred between the two experiments, changed the average age from 67 to 66 without changing the age range. Again, subjects were naïve to the purpose and design of the experiment. NH listeners were familiar with the Dantale II sentences by virtue of participating in Experiment 1, and as noted in Sec. III A 2, HI listeners had experience listening to Dantale II sentences prior to Experiment 1. Due to the limited number of test lists (15) available in the Dantale II corpus, the same lists used in Experiment 1 were also employed in Experiment 2. It is worth mentioning that the corpus was designed for repeated usage (Wagener *et al.*, 2003; see also Sec. III A 1).

3. Procedure and statistical analysis

The procedure of this experiment is the same as in Experiment 1 except for the following. To vary the input SNR, the noise level was adjusted while the speech level was fixed as in the ideal masking conditions of Experiment 1. In the LF and HF conditions, there is no segregation in half of the frequency channels. As the input SNR decreases in the negative range, the sound level of a stimulus in these conditions is dominated by the background noise in the unsegregated frequency range and hence becomes increasingly louder. To ensure that LF and HF stimuli are not too loud for NH listeners who have very low SRTs, the speech level was fixed at a lower volume than in Experiment 1. Despite this change of sound level, all test stimuli were still comfortably audible for NH listeners. Note that this change did not impact HI listeners as the amount of amplification was individually set for them. ANOVA was performed similarly on all the data from NH and HI subjects as in Experiment 1, with within-subject factors of type of processing (LF, HF, or AF) and of type of noise (SSN or CAFE), and a between-subject factor of subject type (NH or HI).

B. Results and discussion

Figure 5 shows the SRT results of all six test conditions in Experiment 2: SSN-LF, SSN-HF, SSN-AF, CAFE-LF, CAFE-HF, and CAFE-AF, for both NH and HI listeners. The ANOVA for NH subjects showed that the main effects of processing type and noise type were significant [$F(2,22)=255.5$, $p<0.001$, and $F(1,11)=231.2$, $p<0.001$, respectively], and there was also a significant interaction between processing type and noise type [$F(2,22)=4.4$, $p<0.05$]. The Bonferroni tests indicated that all NH means were significantly different ($p<0.006$) from one another, except between the SSN-AF and the CAFE-HF condition. For the SSN background, the mean SRT is −15.66 dB in the LF condition, −12.65 dB in the HF condition, and −17.10 dB in the AF condition. The results show that NH listeners perform better when IBM processing is applied in the LF range than in the HF range, and the difference in SRT is approximately 3 dB. This SRT difference is larger than the SRT difference of slightly more than 1 dB reported by Anzalone *et al.* (2006), despite the fact that the boundary separating LFs and HFs is 1.35 kHz in our processing and 1.5 kHz in their processing. Even with the lower frequency boundary we find that, with the same input SNR, the HF condition leaves more noise than the LF condition since the noise energy is distributed mostly in the LF range (see Fig. 4). The discrepancy is likely due to different ways of IBM processing used in the two studies. The AF condition yields the lowest SRT, which is about 1.6 dB lower than in the LF condition.

For the cafeteria background, the mean SRT is −20.37 dB in the LF condition, −17.88 dB in the HF condition, and −23.24 dB in the AF condition. Clearly NH subjects perform better in this background than in SSN, consistent with the results of Experiment 1. Again, NH listeners benefit more from IBM processing at LFs than at HFs and the relative benefit is 2.5 dB. The AF condition also gives the lowest SRT, which is about 2.9 dB lower than in the LF condition. That NH subjects performed better in the AF condition than in the LF condition for both the SSN and cafeteria backgrounds suggest that they do benefit from IBM in the HF range, even though the benefit is not as high as from the LF range.

The ANOVA where both HI and NH subjects were included showed that the main effects of subject type, processing type, and noise type were significant [$F(1,22)=19.1$, $p<0.001$; $F(2,44)=255.4$, $p<0.001$; and $F(1,22)=317.2$, $p<0.001$, respectively], and there were also significant interaction effects between subject type and processing type, subject type and noise type, and processing type and noise type [$F(2,44)=31.2$, $p<0.001$; $F(1,22)=18.3$, $p<0.001$; and $F(2,44)=14.2$, $p<0.001$, respectively], as well as a three-way interaction between subject type, processing type, and noise type [$F(2,44)=5.4$, $p<0.01$]. Table I shows the Fisher LSD *post hoc* tests. As seen in the table, all conditions were significantly different ($p<0.05$) from one another within the NH subjects (conditions {1}–{6} contrasted against each other) and within the HI subjects (conditions {7}–{12} contrasted against each other). However, the differences between NH and HI were insignificant for the conditions of SSN-LF and SSN-AF.

For HI listeners, the mean SRTs for the SSN background are −14.85, −8.49, and −15.96 dB for the LF, HF, and AF conditions, respectively. The SRT advantage of the LF condition over the HF condition is 6.4 dB, whereas the advantage of the AF condition over the LF condition is only

TABLE I. Fisher LSD *post hoc* significance tests for the three-way interaction of subject type, processing type, and noise type. Significance levels above $p > 0.05$ are given in boldface.

Subject type	Processing type	Test condition	{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}	{10}	{11}
NH	SSN-LF	{1} −15.66											
	SSN-HF	{2} −12.65	0.00										
	SSN-AF	{3} −17.10	0.00	0.00									
	CAFE-LF	{4} −20.37	0.00	0.00	0.00								
	CAFE-HF	{5} −17.88	0.00	0.00	**0.07**	0.00							
	CAFE-AF	{6} −23.24	0.00	0.00	0.00	0.00	0.00						
HI	SSN-LF	{7} −14.85	**0.45**	0.05	0.04	0.00	0.01	0.00					
	SSN-HF	{8} −8.49	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
	SSN-AF	{9} −15.96	**0.77**	0.00	**0.29**	0.00	**0.08**	0.00	0.01	0.00			
	CAFE-LF	{10} −18.13	0.03	0.00	**0.34**	0.04	**0.81**	0.00	0.00	0.00	0.00		
	CAFE-HF	{11} −10.05	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	CAFE-AF	{12} −20.96	0.00	0.00	0.00	**0.58**	0.01	0.04	0.00	0.00	0.00	0.00	0.00

1.1 dB. These data are generally comparable with those in Anzalone *et al.* (2006). The results suggest that HI listeners derive considerably more benefit from ideal masking at LFs than at HFs, and the SRT difference is much larger than for NH listeners (see Fig. 5). Although part of the larger gap may be caused by a larger SRT gain (9.2 dB) in HI listeners than that (7.4 dB) in NH listeners due to IBM processing, the fact that the relative advantage of the AF condition over the LF condition for HI listeners is even a little smaller than for NH listeners (1.1 dB versus 1.6 dB) strongly indicates that IBM processing in LF is to a greater extent responsible for the SRT improvement of ideal masking in HI listeners than in NH listeners. In other words, almost all the benefit of IBM can be obtained by IBM only in the LF range. This, of course, is not to say that ideal masking in HF does not improve speech intelligibility compared to no segregation. As illustrated in Fig. 3, IBM processing at all frequencies results in a 9.2 dB SRT improvement compared to no segregation, and the AF condition produces a 7.5 dB relative advantage over the HF condition. This comparison suggests that ideal masking at HFs produces some improvement in speech intelligibility.

For the cafeteria background, the SRTs in the LF, HF, and AF conditions are −18.13, −10.05, and −20.96 dB, respectively (see Fig. 5). The SRT advantage of LF processing over HF processing is 8.1 dB and that of AF over LF is 2.8 dB. These results show a similar pattern as for the SSN background, even though the SRT difference of 2.8 dB between the LF and AF conditions clearly reaches statistical significance (see Table I), and HF processing yields a significant SRT improvement over no segregation as suggested by comparing with the data in Experiment 1. The use of the fluctuating cafeteria background reinforces the conclusion that ideal masking in LF produces a much stronger benefit than that in HF, and this effect is greater in HI listeners than in NH listeners.

The two AF conditions for the SSN and cafeteria backgrounds are the same as the corresponding ideal masking conditions in Experiment 1. The NH performances in Experiment 2 are somewhat better than in Experiment 1. A comparison between Fig. 5 and Fig. 3 shows that the discrepancies are 1.5 dB for SSN and 2.5 dB for cafeteria noise. The only difference in stimuli is the sound level; as pointed out in Sec. IV A 3, the sound level is softer in Experiment 2 than in Experiment 1. For example, at the input SNR of −10 dB, the sound level in Experiment 1 is about 63 dB(A) SPL for the SSN background and 75 dB(A) for the cafeteria background, while the corresponding levels in Experiment 2 are 51 and 51 dB(A), respectively. Studies suggest that softer sound can produce better recognition under certain conditions (Hagerman, 1982; Studebaker *et al.*, 1999). To examine whether the sound volume was a factor in the performance differences, we performed a follow-up experiment with the same pool of the NH listeners who participated in Experiment 2. The follow-up experiment was to simply check subjects' percent correct scores at the sound levels used in the two experiments when the input SNR was fixed at one of the SRTs (alternating between subjects) already obtained in the experiments. The cafeteria background noise was used. The scores are 50.6% with the louder level of Experiment 1 and 58.6% with the softer level of Experiment 2. The 8% difference is statistically significant $[t(11)=3.31,\ p<0.01]$, but unlikely large enough to explain the 2.5 dB SRT difference. Perhaps more important is a learning effect. Unlike HI listeners who were experienced with the Dantale II task, NH listeners used in this investigation had little prior experience with auditory experiments before participating in Experiment 1. When they participated in the second experiment, the familiarity with the Dantale II task acquired during Experiment 1 likely contributed to their better performance. In the predecessor to Dantale II—the Hageman sentence test—Hagerman and Kinnefors (1995) found a training effect of about 0.07 dB per ten sentences, which may explain the differences between Experiments 1 and 2. This interpretation is consistent with the observation that the corresponding performance differences between Experiment 1 and Experiment 2 are smaller for HI listeners; one-third of the mean performance differences is accounted for by the replacement of one HI listener from Experiment 1 to Experiment 2 (see Sec. IV A 2).

V. GENERAL DISCUSSION

The robustness of speech recognition in noise by NH listeners is commonly attributed to the perceptual process of glimpsing, or "listening in the dips," which detects and gathers *T-F* regions of a sound mixture where target speech is relatively stronger compared to interference (Miller and Licklider, 1950; Howard-Jones and Rosen, 1993; Assmann and Summerfield, 2004; Li and Loizou, 2007). As glimpsing involves grouping, this account is closely related to the ASA account that applies to both speech and nonspeech signals (Bregman, 1990). Poorer performance of listeners with hearing loss in fluctuating backgrounds is generally explained as their inability to take advantage of temporal and spectral dips, perhaps caused by reduced frequency selectivity and temporal resolution (Moore, 2007). IBM could be understood as producing glimpses or performing ASA for the listener. The fact that ideal masking also improves intelligibility of NH listeners suggests that even listeners without hearing loss can fail to make full use of the speech information available in a noisy input. The less-than-ideal performance in noisy environments is probably caused by the failure in detecting a glimpse—a *T-F* region with relatively strong target energy—or grouping detected glimpses. This failure becomes more acute with hearing loss. Because ideal masking does an "ideal" job of glimpsing for the auditory system, it helps to nearly equalize the performances of HI and NH listeners (see Fig. 3).

The results of Experiment 1 demonstrate that listeners with or without hearing loss benefit more from IBM processing in the cafeteria background than in the SSN background. The cafeteria background has temporal and spectral modulations, and as a result the amount of informational masking caused by target-masker similarity is expected to be higher than that in SSN. Indeed, some listeners voluntarily commented after the experiment that the conversation in the background distracted their attention, making it harder to concentrate on target utterances. The larger SRT improvement observed for the cafeteria background is thus consistent with the interpretation that ideal masking removes or largely attenuates informational masking (Brungart *et al.*, 2006). In a situation extremely conducive to informational masking, namely, the mixtures of speech utterances of the same talker, Brungart *et al.* (2006) found that the effect of ideal masking is tantamount to a 22–25 dB improvement in input SNR. The 10.5 dB SRT improvement obtained through ideal masking in the cafeteria background, although greater than that obtained in the SSN background, is much smaller than that obtained in mixtures of same-talker utterances. The improvement is also smaller than those obtained in mixtures of different-talker utterances (Chang, 2004), although the gap is not quite as big as in same-talker mixtures. One can therefore expect even larger SRT improvements when interference is one or several competing talkers, a kind of background that produces very large performance gaps between NH and HI listeners as reviewed in Sec. I.

The results of Experiment 2 are on the whole consistent with the related findings of Anzalone *et al.* (2006) even though we used individual gain prescriptions to compensate for listeners' hearing loss. The results are also qualitatively consistent with the findings of Li and Loizou (2007) illustrating that glimpses in the LF to mid-frequency range are more beneficial for speech intelligibility than those in the HF range. However, a few differences between our results and the results of Anzalone *et al.* (2006) are worth noting. First, although considerably smaller than LF processing, there is a benefit from ideal masking in the HF range for HI listeners in our study whereas their study did not show a significant benefit. A possible reason is the individual gain prescription employed in our study that makes segregated speech relatively louder in the HF range than the constant gain applied in their study. Second, we find a relatively greater LF benefit in NH listeners than in their study. The main reason, we believe, is that LF processing removes more background noise than HF processing for a given input SNR. With negative input SNRs (see Fig. 5), the residual noise in the HF condition is in the LF range while that in the LF condition is in the HF range, and the background noises used in our experiments have energy distributed mostly in the LF range, as shown in Fig. 4. This explanation, not considered by Anzalone *et al.*, gives a partial account for the larger LF benefit for listeners with hearing loss. The large SRT gap between LF and HF processing for HI listeners (see Fig. 5), however, cannot be fully explained this way as the gap is substantially larger—to the extent that the SRT performance in LF processing is almost the same as in AF processing. Another likely reason is upward spread of masking (Anzalone *et al.*, 2006) which listeners with sensorineural hearing loss are especially susceptible to (Jerger *et al.*, 1960; Gagne, 1988; Klein *et al.*, 1990). Upward spread of masking is a more prominent factor in the HF condition because of no segregation in the LF range. Also, with more hearing loss at HFs (see Fig. 2), HI listeners are less able to utilize audible HF speech information in recognition compared to NH listeners (Dubno *et al.*, 1989; Ching *et al.*, 1998; Hogan and Turner, 1998). This could also contribute to a steeper performance decline of HF processing relative to AF processing for HI listeners than for NH listeners.

Despite different definitions of IBM, the SRT improvements observed in our study and in Anzalone *et al.* (2006) are very close for the SSN background. It is all the more remarkable considering that their IBM is generated on a sample-by-sample basis while ours is generated on a frame-by-frame basis, which has a drastically lower temporal resolution, and that, in their experiments, IBM-determined gains take the values of 1 and 0.2 while the gains take the values of 1 and 0 in our experiments. The use of two-valued gains is a key similarity between the studies. The most important difference is, of course, that our definition is based on a comparison between target and interference energy and theirs is between target energy and a fixed threshold. The local SNR based IBM is arguably easier to estimate computationally, as many speech segregation algorithms compute binary time-frequency masks by exploring local SNR explicitly or implicitly (Divenyi, 2005; Wang and Brown, 2006). Also, there is little basis in a noisy signal to identify those *T-F* regions of significant target energy where interference is much stronger.

The results from our experiments have major implications for CASA and speech enhancement research aiming to improve speech intelligibility in noisy environments. In addition to affirming the general effectiveness of IBM as a computational goal, our data provide direct evidence that a choice of *LC* at −6 dB for IBM construction, first suggested by Brungart *et al.* (2006), is effective for improving human speech recognition. A comparison between the data of Brungart *et al.* (2006) and ours for the SSN background indicates that the IBM with −6 dB *LC* yields larger SRT improvement than commonly used 0 dB *LC*. Compared to 0 dB *LC*, the choice of −6 dB *LC* retains those *T-F* units where local SNR falls between 0 and −6 dB in ideal masking (see Fig. 1). From the standpoint of SNR, such inclusion will lower the overall SNR of the segregated signal. In other words, if the objective is to improve the SNR of the output signal, the choice of −6 dB *LC* is a poorer one compared to that of 0 dB *LC*. This discussion casts further doubt on the suitability of traditional SNR as a performance metric to evaluate sound separation systems, and at the same time, could shed light on why monaural speech enhancement algorithms often improve SNR but not speech intelligibility (see Sec. I). Another strong implication of our results (see also Anzalone *et al.*, 2006) is that performing speech separation in the LF range is a great deal more important than in the HF range, particularly for improving speech intelligibility of HI listeners.

Our results point to a very promising direction for hearing aid design to improve speech intelligibility in noise of listeners with hearing loss, that is, by designing hearing aids that function in similar ways to IBM. IBM processing improves SRT by a large margin, and HI listeners derive larger benefit than NH listeners. Equally important, the profile of improvement with respect to different kinds of background noise seems to match that of typical hearing impairment. We consider it a highly significant result that ideal masking almost equalizes the intelligibility performances of HI and NH listeners (see Fig. 3). Of course, facing a noisy input IBM cannot be directly constructed and algorithms must be developed to estimate IBM. Encouraging effort has been made in CASA with the explicit goal of IBM estimation (Wang and Brown, 2006), and in limited conditions high-quality estimates are obtainable (see, e.g., Roman *et al.*, 2003). However, computing binary masks close to the IBM in unconstrained acoustic environments remains a major challenge. On the other hand, the extent of intelligibility gain for HI listeners produced by IBM processing much more than fills the SRT gap from NH listeners; Experiment 1 shows a gap of 2.5 dB for the SSN background and a gap of 6.5 dB for the cafeteria background while the ideal masking improvements for HI listeners are 9.2 and 13.8 dB for the two backgrounds, respectively. Hence, perfect IBM estimation is not necessary to bring the performance of HI listeners to the same level as that of NH listeners.

VI. CONCLUSION

The present study was designed to evaluate the impact of IBM on speech intelligibility in noisy backgrounds for both NH and HI listeners. Two experiments were conducted and the main results are summarized below.

- For NH listeners, IBM processing resulted in 7.4 dB SRT reduction for SSN and 10.5 dB reduction for cafeteria noise.
- For HI listeners, IBM processing resulted in 9.2 dB SRT reduction for SSN and 15.6 dB reduction for cafeteria noise.
- After IBM processing, the intelligibility performances for HI listeners and NH listeners were comparable.
- For NH listeners, IBM processing at LFs produced greater SRT reduction than at HFs. The differences were 3 dB for SSN and 2.5 dB for cafeteria noise.
- For HI listeners, IBM processing at LFs produced greater SRT reduction than at HFs. The differences were 5.5 dB for SSN and almost 8 dB for cafeteria noise.

ACKNOWLEDGMENTS

We thank the Associate Editor Ken Grant, and two anonymous reviewers for their helpful comments. The work was conducted while D.W. was a visiting scholar at Oticon A/S. The authors are grateful to M. Schlaikjer, L. Bramsløw, and M. Hartvig, for their assistance in the experiments, and Y. Li for his assistance in figure preparation. D.W. was supported in part by an AFOSR grant (F49620-04-01-0027) and an NSF grant (IIS-0534707).

Alcantara, J. I., Dooley, G., Blamey, P., and Seligman, P. (**1994**). "Preliminary evaluation of a formant enhancement algorithm on the perception of speech in noise for normally hearing listeners," Audiology **33**, 15–27.

Alcantara, J. I., Moore, B. C. J., Kuhnel, V., and Launer, S. (**2003**). "Evaluation of the noise reduction system in a commercial digital hearing aid," Int. J. Audiol. **42**, 34–42.

Anzalone, M. C., Calandruccio, L., Doherty, K. A., and Carney, L. H. (**2006**). "Determination of the potential benefit of time-frequency gain manipulation," Ear Hear. **27**, 480–492.

Assmann, P., and Summerfield, A. Q. (**2004**). "The perception of speech under adverse conditions," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York) pp. 231–308.

Baer, T., Moore, B. C. J., and Gatehouse, S. (**1993**). "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times," J. Rehabil. Res. Dev. **30**, 49–72.

Beck, S., and Zacharov, N. (**2006**). *Perceptual Audio Evaluation: Theory, Method and Application* (Wiley, Chichester, NY).

Benesty, J., Makino, S., and Chen, J., eds. (**2005**). *Speech Enhancement* (Springer, New York).

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Brungart, D. S. (**2001**). "Information and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D., Chang, P. S., Simpson, B. D., and Wang, D. L. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.

Bunnell, H. T. (**1990**). "On enhancement of spectral contrast in speech for hearing-impaired listeners," J. Acoust. Soc. Am. **88**, 2546–2556.

Carhart, R. C., and Tillman, T. W. (**1970**). "Interaction of competing speech signals with hearing losses," Arch. Otolaryngol. **91**, 273–279.

Chang, P. (**2004**). "Exploration of behavioral, physiological, and computational approaches to auditory scene analysis," M.S. thesis, The Ohio State University Department of Computer Science and Engineering, Columbus, OH; http://www.cse.ohio-state.edu/pnl/theses.html (Last viewed September 2008).

Ching, T. Y. C., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.
Dillon, H. (**2001**). *Hearing Aids* (Thieme, New York).
Divenyi, P., ed. (**2005**). *Speech Separation by Humans and Machines* (Kluwer Academic, Norwell, MA).
Drullman, R. (**1995**). "Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level," J. Acoust. Soc. Am. **98**, 1796–1798.
Dubno, J. R., Dirks, D. D., and Ellison, D. E. (**1989**). "Stop-consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss. I: The contribution of selected frequency regions," J. Acoust. Soc. Am. **85**, 347–354.
Edwards, B. (**2004**). "Hearing aids and hearing impairment," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York).
Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Hear. Res. **38**, 222–233.
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.
Gagne, J.-P. (**1988**). "Excess masking among listeners with a sensorineural hearing loss," J. Acoust. Soc. Am. **83**, 2311–2321.
Greenberg, J. E., and Zurek, P. M. (**1992**). "Evaluation of an adaptive beamforming method for hearing aids," J. Acoust. Soc. Am. **91**, 1662–1676.
Hagerman, B. (**1982**). "Sentences for testing speech intelligibility in noise," Scand. Audiol. **11**, 79–87.
Hagerman, B., and Kinnefors, C. (**1995**). "Efficient adaptive methods for measurements of speech reception thresholds in quiet and in noise," Scand. Audiol. **24**, 71–77.
Hansen, M., and Ludvigsen, C. (**2001**). "Dantale II—Danske Hagermann sætninger (Dantale II—Danish Hagermann sentences)," Danish Speech Audiometry Materials (Danske Taleaudiomaterialer), Værløse, Denmark.
Hogan, C. A., and Turner, C. W. (**1998**). "High-frequency audibility: Benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.
Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomudulated glimpsing in 'checkerboard' noise," J. Acoust. Soc. Am. **93**, 2915–2922.
Hygge, S., Ronnberg, J., Larsby, B., and Arlinger, S. (**1992**). "Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds," J. Speech Hear. Res. **35**, 208–215.
Jerger, J. F., Tillman, T. W., and Peterson, J. L. (**1960**). "Masking by octave bands of noise in normal and impaired ears," J. Acoust. Soc. Am. **32**, 385–390.
Johannesson, R. B. (**2006**). "Output SNR measurement method," Report No. 052-08-04, Oticon Research Centre Eriksholm, Snekkersten, Denmark.
Kates, J. M., and Weiss, M. R. (**1996**). "A comparison of hearing-aid array-processing techniques," J. Acoust. Soc. Am. **99**, 3138–3148.
Klein, A. J., Mills, J. H., and Adkins, W. Y. (**1990**). "Upward spread of masking, hearing loss, and speech recognition in young and elderly listeners," J. Acoust. Soc. Am. **87**, 1266–1271.
Levitt, H. (**2001**). "Noise reduction in hearing aids: A review," J. Rehabil. Res. Dev. **38**, 111–121.
Li, N., and Loizou, P. C. (**2007**). "Factors influencing glimpsing of speech in noise," J. Acoust. Soc. Am. **122**, 1165–1172.
Li, N., and Loizou, P. C. (**2008a**). "Effect of spectral resolution on the intelligibility of ideal binary masked speech," J. Acoust. Soc. Am. **123**, EL59–EL64.
Li, N., and Loizou, P. C. (**2008b**). "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," J. Acoust. Soc. Am. **123**, 1673–1682.
Li, Y., and Wang, D. L. (**2009**). "On the optimality of ideal binary time-frequency masks," Speech Commun. **51**, 230–239.
Lim, J., ed. (**1983**). *Speech Enhancement* (Prentice-Hall, Englewood Cliffs, NJ).
Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.
Moore, B. C. J. (**2003a**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego, CA).
Moore, B. C. J. (**2003b**). "Speech processing for the hearing-impaired: Successes, failures, and implications for speech mechanisms," Speech Commun. **41**, 81–91.
Moore, B. C. J. (**2007**). *Cochlear Hearing Loss*, 2nd ed. (Wiley, Chichester, UK).
Nilsson, M., Soli, S., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.
Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.
Plomp, R. (**1994**). "Noise, amplification, and compression: Considerations of three main issues in hearing aid design," Ear Hear. **15**, 2–12.
Rabiner, L. R., and Juang, B. H. (**1993**). *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ).
Ricketts, T., and Hornsby, B. W. (**2003**). "Distance and reverberation effects on directional benefit," Ear Hear. **24**, 472–484.
Roman, N., Wang, D. L., and Brown, G. J. (**2003**). "Speech segregation based on sound localization," J. Acoust. Soc. Am. **114**, 2236–2252.
Schum, D. J. (**2003**). "Noise-reduction circuitry in hearing aids, II: Goals and current strategies," Hear. J. **56**, 32–41.
Simpson, A. M., Moore, B. C. J., and Glasberg, B. R. (**1990**). "Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners," Acta Oto-Laryngol. **469**, 101–107.
StatSoft, Inc. (**2007**). STATISTICA (data analysis software system), version 7, http://www.statsoft.com (Last viewed February 2008).
Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.
Vestergaard, M. (**1998**). "The Eriksholm CD 01: Speech signals in various acoustical environments," Report No. 050-08-01, Oticon Research Centre Eriksholm, Snekkersten, Denmark.
Wagener, K., Josvassen, J. L., and Ardenkjær, R. (**2003**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol. **42**, 10–17.
Wang, D. L. (**2005**). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Norwell, MA), pp. 181–197.
Wang, D. L., and Brown, G. J., eds. (**2006**). *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications* (Wiley, Hoboken, NJ/IEEE, New York).

Perception of allophonic cues to English word boundaries by Japanese second language learners of English

Kikuyo Ito[a] and Winifred Strange
Ph.D. Program in Speech-Language-Hearing Sciences, City University of New York—Graduate School and University Center, 365 Fifth Avenue, New York, New York 10016-4309

(Received 30 April 2008; revised 10 December 2008; accepted 23 January 2009)

Perception of stop aspiration and glottal stop allophonic cues for word juncture in English by Japanese second language (L2) learners of English was examined, extending a study of Spanish L2 learners [Altenberg, E. P. (2005). Second Lang. Res. **21**, 325–358]. Thirty Japanese listeners ranging in length of residence (LOR) in the United States (2 weeks–12 years) were tested on 42 contrasting pairs (e.g., aspiration: *keeps talking* vs *keep stalking*, glottal stop: *a nice man* vs *an ice man*, and double cues: *grape in* vs *grey pin*). Phrases were presented in randomly ordered lists and subjects responded in a two-choice identification task followed by a phrase familiarity test. The Japanese listeners performed more poorly than an American English-speaking control group ($N=10$), especially on aspiration pairs. Aspiration pairs were differentiated significantly less well (73% correct) by Japanese listeners than were glottal stop pairs (91% correct) and double cue pairs (94% correct); response biases predicted from relative familiarity of phrases were evident only for aspiration pairs. Performance correlated with LOR and suggested that aspiration cues take more immersion experience to learn than glottal stop cues. The patterns of errors were similar, but not identical, to Altenberg's Spanish data.
 [DOI: 10.1121/1.3082103]

PACS number(s): 43.71.Hw, 43.71.Es [PEI] Pages: 2348–2360

I. INTRODUCTION

When processing running speech, listeners have to segment continuous streams of sound into recognizable units such as words and phrases. Identifying word/phrase boundaries is a necessary stage for auditory language processing because there are no reliable pauses between words in continuous speech. The process of word segmentation in speech seems to be so automatic and effortless for first language (L1) listeners that, more often than not, they are not even aware of doing it. On the other hand, everybody has experience of not being able to figure out where the words start and end when listening to unfamiliar foreign speech, and even with learning experience, correctly segmenting a stream of non-native speech into words and phrases often requires extensive exposure to the language and a considerable amount of practice.

Various kinds of information may be involved in the process of word segmentation: lexical/semantic cues (e.g., Gow and Gordon, 1995), syntactic/morphological cues (e.g., Sanders and Neville, 2000), phonotactic cues (e.g., McQueen, 1998), prosodic/rhythmic cues (e.g., Cutler and Norris, 1988), and allophonic acoustic-phonetic cues (e.g., Nakatani and Dukes, 1977). Native listeners are likely to utilize all these multiple cues available in natural speech for efficient word segmentation, and several cue-trading segmentation studies by native English listeners have been documented. For example, Gow and Gordon (1995) examined the interaction of acoustic-phonetic and lexical information on word segmentation and found that lexical contexts induced simultaneous access to words other than those intended by speakers (e.g., accessing *tulips* in addition to *two lips* when listening to *two lips*). Mattys *et al.* (2005) investigated how word stress cues interacted with conflicting other cues and found that stress information became a strong segmentation cue when acoustic-phonetic, phonotactic, or lexical information was not available. Their results also indicated that acoustic-phonetic and phonotactic information were secondary to lexical knowledge, suggesting that higher-order information outweighs sublexical cues. Mattys and Melhorn (2007) also examined the effects of sentential context, lexical knowledge, and acoustic-phonetic cues in connected speech and found compensatory mechanisms between them: The more ambiguous the acoustic-phonetic cues, the larger the effect of both sentential and lexical context on segmentation when listening to near-homophonous phrases (e.g., /plʌmpaɪ/for *plum pie* or *plump eye*). They also found that lexical effects reduced sensitivity to the acoustic-phonetic cues (e.g., favoring *scum pie* over **scump eye* even when it was acoustically more like **scump eye*), highlighting the graded, interactive, and flexible nature of multi-cue segmentation. The present study examined how non-native listeners process acoustic-phonetic segmentation cues present in L2 speech to perceive word boundaries when other types of cues are not available.

In a series of studies on the use of multiple segmentation cues, Sanders and co-workers compared strategies in segmenting English input used by monolingual native speakers of English (Sanders and Neville, 2000, 2003a) and by late L2 learners of English (Sanders *et al.*, 2002; Sanders and Neville, 2003b). The studies used behavioral as well as event-

[a]Electronic mail: kikuyoito@hotmail.com

related potential (ERP) measurements and focused on the interaction of lexical, syntactic, and stress-pattern cues. The behavioral results for English listeners confirmed that they exploited these cues in combination (Sanders and Neville, 2000) whereas native Japanese and native Spanish late learners did not use syntactic information to the same extent as native listeners did, although they exploited lexical information to a similar extent (Sanders *et al.*, 2002). The results also indicated that Japanese late learners as well as Japanese near monolinguals used the typical English stress cues for segmentation better than Spanish late learners did, suggesting that the extent to which a cue was used depended on the characteristics of their native languages. In the ERP experiments, native English listeners showed strong word-onset effects, suggesting that neither semantic nor syntactic information is necessary for word-onset identification (Sanders and Neville, 2003a). Japanese late learners' ERP data, on the other hand, only showed semantic processing effects without showing the word-onset effects nor the syntactic processing effects (Sanders and Neville, 2003b). These results indicated that non-native listeners were not processing the acoustic differences in the same way as native listeners.

While the research on the use of multiple segmentation cues has increasingly drawn attention recently, studies on individual segmentation cues have not been as well documented to date. The bulk of segmentation research dealing with individual cues has concentrated on examining the use of rhythmic information, especially in the area of cross-language research. (e.g., Mehler *et al.*, 1981; Cutler and Norris, 1988; Cutler and Butterfield, 1992; Otake *et al.*, 1993; Cutler *et al.*, 1986, 1992; Cutler and Otake, 1994; Bradley *et al.*, 1993; Goetry and Kolinsky, 2000). Although the research on rhythmic cues has made a significant contribution to our understanding of native and non-native word segmentation, studies on other individual cues have not been well explored yet. The present study contributes by investigating the use of allophonic acoustic-phonetic segmentation cues (*allophonic cues*, hereafter) for word juncture in American English (AE, hereafter): (a) presence/absence of aspirated voiceless stop consonants (aspirated in word-initial stops) and (b) presence/absence of glottal stops as cues to word onset (present before word-initial vowels).[1]

In studies on the use of allophonic segmentation cues by native listeners, researchers have found that the occurrence of allophones of the same phoneme can signal segmentation. Christie (1974) examined English listeners' perception of syllable boundaries by using synthetic stimuli and found that the aspiration of /t/ affected the perception of the boundary location whereas formant transitions had little or no effect on the boundary. Nakatani and Dukes (1977) examined English listeners' use of allophonic variations as segmentation cues, using spliced speech of near-homophonous phrases such as *no notion/known ocean* and *play taught/plate ought*. The authors found that glottal stops or laryngealization (creaky voice) at the onset of word-initial vowels, aspiration of voiceless stops, and syllable-initial vs -final /l/ and /r/ allophones were strong segmentation cues for AE listeners. They also found that duration differences of words caused by presence/absence of voiceless word-final consonants (e.g., *plate* in *plate ought* being longer than *play* in *play taught*) were not useful as a segmentation cue for native listeners. Church (1987) also argued from a computational linguistic viewpoint that the information provided by allophonic variations is useful for segmentation. He illustrated that the systematic distribution of allophonic variations, such as word-initial aspiration of stops, stop release, flapping, and palatalization in English, facilitates parsing speech segments by determining the location of syllables and stress.

Cross-language research on word segmentation dealing with allophonic cues is very limited. Altenberg (2005) investigated whether Spanish L2 learners of English could use English aspiration cues and glottal stop cues for segmentation in the same way as native listeners did, using an adapted version of the design by Nakatani and Dukes (1977). Participants were presented speech stimuli, such as *keeps talking* and *a nice man*, and were asked to indicate whether they heard *keep stalking* or *keeps talking*, and *a nice man* or *an ice man*. In these examples, the presence of aspiration in /t/ in *keeps talking*, which is absent in *stalking*, and the presence of glottal stop or creaky voice preceding *ice* in *an ice man*, which is absent in *nice*, provide strong segmentation cues for native speakers of English (Nakatani and Dukes, 1977; Altenberg, 2005). Altenberg (2005) found that Spanish listeners perceived the word boundaries significantly less accurately than English listeners did. She also found that Spanish listeners' performance on the aspiration pairs was significantly worse than that on the glottal stop pairs, while no differences were seen in the performance of native English speakers. Altenberg (2005) concluded that Spanish L2 learners of English were unable to fully use these allophonic cues available to native English listeners and suggested that various factors, including L1 transfer and markedness, may be responsible for greater ease or difficulty of utilizing particular L2 word segmentation cues. Further research is needed to see if the observed relative difficulties generalize to listeners of other languages and to explore the factors that may cause ease or difficulty in processing L2 allophonic segmentation cues.

The present study is an extension of the Altenberg study (2005) of aspiration and glottal stop segmentation cues. Japanese L2 learners of English served as experimental subjects with native English speakers as a control group. The distribution and the extent of aspiration of voiceless stops in Japanese are different from those in English and also different from those in Spanish. Japanese voiceless stops /p/, /t/, and /k/ are weakly aspirated in word-initial position but unaspirated elsewhere (Hattori, 1951; Vance, 1987). According to Homma (1980), the voice onset time (VOT) of Japanese word-initial /t/ is approximately 25 ms. Recently, Riney *et al.* (2007) measured VOT values of Japanese /p/, /t/, and /k/ produced in several consonant + vowel (CV) contexts and compared them with those of English /p/, /t/, and /k/ produced in equivalent conditions by monolingual AE speakers. The mean values of Japanese /p/, /t/, and /k/ were 30.0, 28.5, and 56.7 ms, respectively, whereas the corresponding AE VOT values were 78.2, 94.8, and 102.8 ms. As for the VOT values of Spanish initial voiceless stops, the Latin American Spanish data by Williams (1977) seem to be the closest approximation of the native VOT values for Alten-

berg's (2005) Spanish participants, who were primarily Dominicans, Mexicans, and Panamanians. In Williams' data (1977), the average VOT values of /p/, /t/, and /k/ in Guatemalan, Venezuelan, and Peruvian Spanish ranged from 9.8 to 15.2 ms, 10.3 to 20.6 ms, and 25.7 to 32.6 ms, respectively. Thus, since aspiration of voiceless stops in initial position appears to be longest in AE, shortest in Spanish, and of intermediate duration in Japanese, we can assume that allophonic differences between initial and non-initial voiceless stops are greatest for AE, least for Spanish, and intermediate for Japanese, assuming that values for noninitial stops are near zero VOT for all three languages.

Glottal stops in Japanese frequently occur in utterance-final position in emphatic speech (Bloch, 1950; Vance, 1987), and some researchers further argue that any utterance-final short vowel in Japanese is followed by a glottal stop (Kawakami, 1977; Vance, 1987). A glottal stop in Japanese generally appears before an utterance-initial vowel in emphatic speech (Vance, 1987) as well and it may appear even before a word-initial or phrase-initial vowel (Kawakami, 1977). Thus, in Japanese, glottal stops appear to assume a role as an utterance-, phrase-, or word-boundary marker. A glottal stop in Japanese can be realized as an allophonic variation of the moraic obstruent /Q/ (i.e., the first half of geminate consonants) as well. However, it cannot be confusable with the boundary-marking glottal stops because the moraic obstruent /Q/ cannot occur utterance finally nor immediately before a vowel. In Spanish, as with Japanese, a glottal stop can be inserted before a word-initial vowel, and after a word-final vowel in emphatic speech (Stockwell, *et al.*, 1956).

Thus, if ease or difficulty in using non-native segmentation cues is based on their use in the native language (i.e., L1 transfer), Japanese listeners in the present study would be expected to show better perception of aspiration cues than Spanish listeners (Altenberg, 2005), while showing similar performance on glottal stop cues. By examining the perception of Japanese listeners using the same lexical items as the Altenberg study (2005), the present study contributes new information on the use of allophonic cues for L2 word segmentation.

In this study a new native speaker of AE produced the materials, using the same instructions as were used in the original experiment. A control group of AE listeners was included to verify that the materials contained sufficient segmentation cues for natives to perform with near ceiling as did Altenberg's control group (2005). Detailed acoustic analysis was also conducted to document the multiple parameters that distinguished the stimulus pairs. Two additional testing procedures were added: (a) listeners' confidence ratings of their response choices, and (b) a follow-up familiarity rating test. The latter was used to establish whether the familiarity of members of the pairs differed systematically with respect to the presence/absence of aspiration/glottal stop cues to segmentation and to help in the interpretation of the results.[2]

The research questions in the present study were as follows.

(1) Is Japanese L2 listeners' use of English allophonic cues for word segmentation poorer than that of AE native listeners? It was expected that AE listeners would perform at ceiling, whereas Japanese listeners would perform significantly worse on both kinds of segmentation cues. However, in addition to group differences, establishing AE listeners' performance levels on the same materials allowed us to compare individual Japanese participants' performance against a native-speaker level of accuracy as a function of immersion experience.
(2) Are aspiration cues more difficult for Japanese listeners to utilize than glottal stop cues? Given the distribution of such cues in Japanese, we expected that performance on aspiration cues would be significantly poorer than performance on glottal stop cues. In addition, performance on presence vs absence of the cues was compared, given Altenberg's finding (2005) that the presence of aspiration and glottal stop cues provided more effective information than their absence.
(3) Does Japanese listeners' use of English allophonic cues improve with more experience in an English-speaking environment? Correlational analyses of immersion experience and performance on each type of segmentation cues were also employed to explore whether some cues take more immersion experience to learn than others.

II. METHODS

A. Participants

Thirty adult native speakers of Japanese living in the United States and a control group of ten adult native speakers of AE participated in the experiment. All participants passed a hearing screening (ANSI standards 25 dB HL at 500, 1 000, 2 000, and 4 000 Hz). Information about the Japanese participants, such as age, gender, length of residence (LOR) in English-speaking countries, which includes stays in the United States, Canada, United Kingdom, or Australia,[3] age of arrival (AOA) in English-speaking countries, and educational experience in English-speaking countries (immersion experience), was collected by means of a questionnaire. Since none of the Japanese participants came to English-speaking countries earlier than 15 years of age (AOA: 15–45 years of age), all Japanese participants were considered late learners of English. The information for the AE participants was also collected via a questionnaire. All AE participants had had experience in learning at least one of the following languages at school: Spanish, French, Italian, or Latin. None of them, however, had learned Japanese or any Asian languages in their lives. The age range and gender proportion of the AE group were matched to those of the Japanese group as closely as possible. Table I shows the information about the Japanese participants and the AE participants.

B. Stimulus materials

Eighty-four naturally produced stimuli that constitute forty-two stimulus pairs were recorded. The stimuli were identical to those used in the Altenberg (2005) study but were produced by a different speaker: a phonetically trained

TABLE I. Biographical information for AE participants and Japanese participants (age, gender, LOR, AOA, and immersion experience).

	American participants	Japanese participants
Sample size	10 (9 females, 1 male)	30 (26 females, 4 males)
Mean age (range)	31.2 years (26–38 yrs)	30.7 years (20–46 years)
Mean LOR[a] (range)		4.0 years (2 weeks–12 years)
Mean AOA[b] (range)		25.0 years (15–45 years)
Mean immersion experience[c] (range)		2.5 years (0–10 years)

[a]LOR=length of residence in English-speaking countries.
[b]AOA=age of arrival in English-speaking countries.
[c]Immersion experience=educational experience in English-speaking countries (includes experience at language school).

female native speaker of AE with a New York dialect. The stimuli were produced in the carrier phrase "Say__again." Following the procedure of the Altenberg study (2005), the speaker was instructed to speak naturally but at a slightly slower rate than usual to minimize the likelihood of listeners' performance errors (p. 336). Altenberg (2005) noted that care was taken to minimize the influence of lexical familiarity. The stimuli were selected from a larger set of 168 phrases (84 pairs) based on naturalness ratings by 7 native speakers of English "such that the naturalness rating of one member of the set (e.g., *keep stalking*) would be as close as possible to the naturalness rating of the other member of the set (e.g., *keeps talking*)" (p. 336). All stimuli were digitally recorded in a sound-attenuated room at 22.050 kHz, monaural, with 16 bit resolution, using a microphone (SHURE SM 48) fed through a preamplifier (Earthworks LAB 101) into the computer analog to digital sound card using SOUND FORGE 4.5 software.

The stimuli are subdivided into three types: 18 Aspiration stimulus pairs (36 phrases) in which presence or absence of aspiration of a voiceless-stop at a word boundary signals the boundary, 18 Glottal stop stimulus pairs (36 phrases) in which presence or absence of glottal stop or creaky voice signals the boundary, and 6 Double cue stimulus pairs (12 phrases) in which both presence or absence of aspiration and of glottal stop can be used as cues for segmentation.

Aspiration stimuli are further broken down into three groups: the stimuli that have a vowel before the /s/ at the juncture followed by a voiceless stop (VsC group; e.g., *loose tops/Lou stops*), the stimuli that have a consonant before the /s/ followed by a voiceless stop (CsC group; e.g., *keeps talking/keep stalking*), and the stimuli that have a consonant before the /s/ followed by a voiceless stop plus another consonant (CsCC group; e.g., *cook's prints/cook sprints*). Glottal stop stimuli are also broken down into three groups: the stimuli that have a nasal consonant at the juncture (Nasal group; e.g., *an itch/a niche*[4]), the stimuli that have an obstruent at the juncture (Obstruent group; e.g., *wife ill/why fill*), and the stimuli that have a liquid at the juncture (Liquid group; e.g., *I'll earn/I learn*). The stimuli with an aspiration cue are referred to as *Positive Aspiration stimuli* and those with a glottal stop cue as *Positive Glottal stop stimuli*. The stimuli without aspiration and those without glottal stop are referred to as *Negative Aspiration stimuli* and *Negative Glottal stop stimuli*, respectively. Double cue stimuli do not have subcategories. The items of each pair are either one of the following two types: *Aspiration-negative/Glottal-positive* (hereafter *Asp−/Glot+*) *stimuli*, which contain the glottal stop cue but no aspiration cue (e.g., *grape in*); or *Aspiration-positive/Glottal-negative* (hereafter *Asp+/Glot−*) *stimuli*, which contain the aspiration cue but no glottal stop cue (e.g., *grey pin*).

In addition to the 42 stimulus pairs, 6 extra stimulus pairs (12 stimuli) for familiarization trials were recorded by the same speaker. These stimulus pairs were taken from those used in Nakatani and Dukes, 1977, except for one stimulus pair (*ice cream/I scream*). All stimulus pairs, including those used in the familiarization trials, are presented in the Appendix.

C. Acoustic analysis

An acoustic analysis was carried out for all 84 items to detect any consistent acoustic cues that signal the word juncture in the stimuli, using the computer software MULTI-SPEECH. The critical area of the target phrase, including pre-boundary and post-boundary segments, was selected and displayed on the screen in waveforms and wideband spectrograms that were temporally aligned. The selected areas contained four to five phonemes (e.g., /ipstɔ/ in *keeps talking* and *keep stalking*). The analysis size of the spectrogram was set at 100 points (323.00 Hz) with a 0–5000 Hz display in Hamming window. The durations of distinct acoustic segments, such as silence period, frication, VOT, stop burst, and nasal murmur, were measured by placing cursors in reference to both waveform and spectrogram. Where successive segments were difficult to divide further, as in a sequence of an approximant and a vowel, it was treated as one unit to avoid inaccurate segmentation.

1. Aspiration stimuli

The left panel of Fig. 1 shows the durations of acoustic segments observed in Aspiration stimuli that were averaged across the subcategorized groups with positive and negative distinctions shown separately. The primary aspiration cue was quantified by measuring the duration of VOT, which was seen as noise between the onset of the stop burst and the first striation of the following vowel in the spectrogram. As predicted, the VOTs were distinctively longer in Positive (range 40–114 ms) stimuli than in Negative (range=0–53 ms)

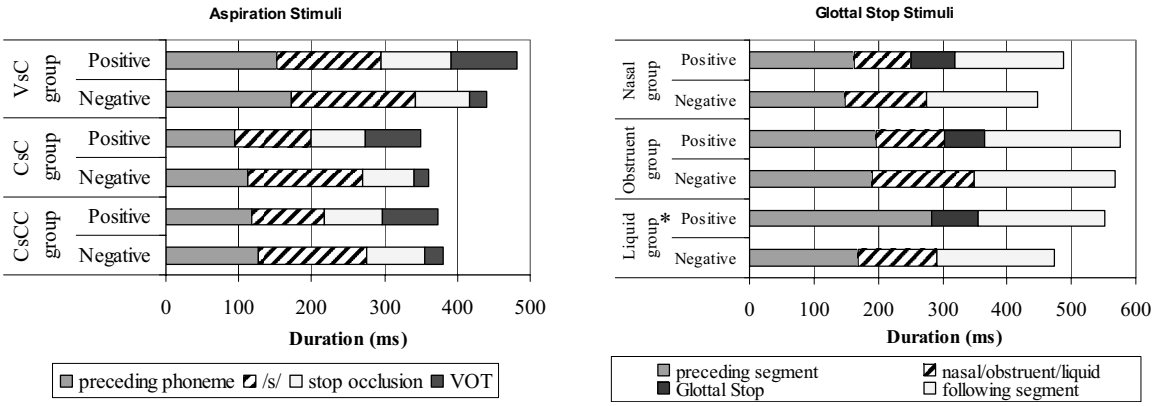


FIG. 1. Mean durations of acoustic segments: Aspiration stimuli (left panel) and Glottal stop stimuli (right panel). *In the positive stimuli of the Liquid group, the liquid segment was not distinguishable from the preceding segment. For this subgroup, preceding segment includes both the preceding segment as well as the liquid segment.

stimuli in all three subcategories, and the positive items were always longer than the negative counterparts (differences ranging from 22 to 95 ms). This difference was highly significant by a paired sample *t*-test (t [17]=12.17, $p<0.001$). In addition, the /s/ frication durations were noticeably shorter in the Positive than in the Negative stimuli (t [17]=5.92, $p<0.001$) and the preceding phonemes were slightly shorter in Positive stimuli than in Negative stimuli in all three subcategories (t [17]=2.46, $0.01<p<0.05$).

2. Glottal stop stimuli

The right panel of Fig. 1 shows the mean durations of acoustic segments in Glottal stop stimuli categorized in the same way as those in Aspiration stimuli. The most distinctive acoustic cue of Glottal stop stimuli, identified as glottal stop or creaky voice, was characterized by the silent gap or aperiodic sound preceding the word-initial vowel. None of the negative stimuli contained these cues. In Positive stimuli, there were two tokens, *an ice man* and *save amps*, that did not contain the glottal stop/creaky voice. They sounded potentially confusable with the negative counterparts, *a nice man* and *say vamps*, to the first author who was a native speaker of Japanese. However, two native speakers of AE judged them as the correct phrases and as sounding natural. Thus, these two tokens were kept as stimuli to see if native listeners were still able to correctly perceive them. The performance on these two tokens by the Japanese group and by the AE group is discussed in Sec. III B. Durational differences between positive and negative differences, including the two aberrant tokens, were statistically significant (t [17] =8.60, $p<0.001$).

It was also observed that the nasal murmurs in the Nasal group and the /f/ or /v/ frication in the Obstruent group were significantly shorter in Positive stimuli than in Negative stimuli (t [11]=3.15, $p<0.01$).[5] Furthermore, in the Liquid group, there was an obvious difference between Positive and Negative stimuli. While the liquid portion and the preceding vocalic part were distinguishable in spectrograms in negative stimuli (e.g., /l/ in *I learn*), they were not separable in the positive group (e.g., /l/ in *I'll earn*), due to their pronunciation as postvocalic allophones that are continuous with preceding vowel segments (vocalic [ɚ] and velarized [ɫ] in the above example). As a result, the "preceding segment" of positive stimuli in the Liquid group in the right panel of Fig. 1 includes the liquid part as well.

3. Double cue stimuli

As mentioned earlier, in Double cue stimuli, one type of item is characterized by the absence of aspiration and presence of glottal stop (*Asp−/Glot+ stimuli*), and the other by the presence of aspiration and absence of glottal stop (*Asp+/Glot− stimuli*). The acoustic analysis not only confirmed these acoustic characteristics but also suggested additional acoustic cues for word segmentation.

In Asp−/Glot+ stimuli, positive VOTs were not seen in either of the stimuli, and all stimuli contained silence or random noise that indicated glottal stop or creaky voice before the word-initial vowel with one exception, *light old*. In this particular token, there was a 108 ms of silence before the vowel, but the silence originated as a stop occlusion so it was not possible to judge whether a glottal stop was included in the silence or not. In addition to these expected acoustic cues, in Asp−/Glot+ stimuli with a bilabial or velar stop (i.e., *grape in, weep at, make art, like old*), 38–46 ms stop bursts were observed between the stop occlusion and the glottal stop/creaky voice, which can be a strong word juncture cue. In Asp+/Glot− stimuli, 105–122 ms VOTs, which were totally absent in Asp−/Glot+ stimuli, were observed, and no signs of glottal stop/creaky voice were seen. Additional possible juncture cues included shorter stop silences in Asp−/Glot+ stimuli (50–113 ms) than in Asp−/Glot+ stimuli (83–129 ms), except for the *light old/lie told* pair. Moreover, the segments preceding the stop occlusion in Asp −/Glot+ stimuli (117–339 ms) were always shorter than those in corresponding Asp−/Glot+ stimuli (152–410 ms).

4. Summary

Table II summarizes the acoustic cues confirmed by the acoustic analysis. The acoustic analysis revealed that the difference in VOT values was present as the most distinctive acoustic cue in Aspiration stimuli and further indicated that the duration of the /s/ frication, and possibly that of the preceding phoneme, may also signal the word juncture. For

TABLE II. Summary of acoustic cues for word juncture.

Stimulus type	Main cues	Secondary cues	Other cues
Aspiration stimuli			
Positive stimuli (e.g., *loose tops*)	Longer VOT	Shorter /s/ friction	Shorter preceding vocalic segment
Negative stimuli (e.g., *Lou stops*)	Shorter VOT	Longer /s/ frication	Longer preceding vocalic segment
Glottal stop stimuli			
Positive stimuli (e.g., *an itch, wife ill, seal of*)	Presence of glottal stop	• Shorter nasal murmur or /f/, /v/ frication • Post-vocalic /l/ or /r/	
Negative stimuli (e.g., *a niche, why fill, sea love*)	Absence of glottal stop	• Longer nasal murmur or /f/, /v/ frication • Pre-vocalic /l/ or /r/	
Double cue stimuli			
Asp−/Glot+ (e.g., *grape in*)	• Absence of +VOT • Presence of glottal stop	Stop burst of /p/ or /k/ (but not /t/)	• Shorter stop silence • Shorter preceding segment
Asp+/Glot− (e.g., *grey pin*)	• Presence of +VOT • Absence of glottal stop		• Longer stop silence • Longer preceding segment

Glottal stop stimuli, the clear difference of presence or absence of the glottal stop/creaky voice was the most reliable acoustic cue. The duration of the nasal murmur, or the /f/ or /v/ frication, was an additional cue for the word juncture. In the case of the Liquid group, the allophonic change of /r/ or /l/ suggested a strong cue for word juncture. For Double cue stimuli, the combination of the presence or absence of long VOT and glottal stop/creaky voice indicated strong juncture cues. In addition, the acoustic analysis suggested that the stop burst of /p/ or /k/ (but not /t/), the duration of the stop silence, and that of the preceding segment could also be juncture cues. The longer duration of frication, nasal murmur, and stop silence at word-initial position observed across all stimulus types correspond with the findings of the lengthening of word-initial consonants in early studies (e.g., Oller, 1973; Furgeron and Keating, 1997).

D. Procedures

The participants were tested individually in a sound-attenuated room in the Speech-Language-Hearing Sciences Department at the CUNY Graduate Center, using computer software, PARADIGM ID (written by Tagliaferri) in which participants responded by clicking on one of two response alternatives printed on the screen.[6] The experiment consisted of the main experiment and the familiarity questionnaire.

1. Main experiment

The stimuli were presented in random order, except that the members of each stimulus pair were separated by at least 15 other stimuli. The stimuli were presented binaurally through headphones (Telephonics TDH 39). The sound level was set in such a way that a 1000-Hz pure tone with the same amplitude as the median rms value of all stimuli was presented at 70 dB SPL. For each trial, the printed stimulus pair appeared on the computer screen approximately 1 s after the presentation of the auditory stimulus. One member of each stimulus pair on the screen always matched the auditory stimulus. Each pair was placed side by side and the position of the correct answer was randomized. Participants were informed by written instructions on the screen that they would hear two to three English words embedded in the sentence "Say__ again," and that they would be asked to click on one of the written alternatives that represented the speech stimulus. After choosing one member of the pair, they were asked to indicate how confident they were in their answer by clicking on one of the seven boxes corresponding to the seven levels of confidence rate (1=*not sure* to 7=*very confident*). They were specifically instructed that though some of the target phrases might sound awkward or unfamiliar, their choice should be based only on what they heard, not on the familiarity of the words. Participants responded to six sample stimulus pairs in the familiarization test before doing the main experiment.

2. Familiarity questionnaire

The main experiment was followed by the familiarity questionnaire that asked the participants whether one of each stimulus pair was more familiar than the other. The written stimulus pairs were presented on the screen without speech stimuli, and participants were asked to click on the member of the pair that was more familiar to them, or to click on *Same* when the two alternatives were equally familiar to them.

III. RESULTS

This section reports the results in the following order: (1) Japanese listeners' performance vs AE listeners as a group, (2) item analysis, and (3) correlation between Japanese listeners' LOR and performance. In the first part, the overall performance by the Japanese group, including their

TABLE III. Mean percent correct for each stimulus type (positive vs negative): japanese and AE subjects.

	American subjects (n=10)			Japanese subjects (n=30)		
Stimulus type	Mean	SD[a]	SE[b]	Mean	SD	SE
Aspiration (n=36) (e.g., Lou stops/loose tops)	95.8	4.4	1.46	73.1	9.4	1.74
Positive aspiration (n=18; e.g., loose tops)	98.3	3.7	1.25	73.5	15.5	2.88
Negative aspiration (n=18; e.g., Lou stops)	93.3	9.0	3.00	72.8	13.0	2.41
Glottal stop (n=36) (e.g., a niche/an itch)	96.9	1.6	0.53	91.3	6.2	1.15
Positive glottal stop (n=18; e.g., an itch)	98.9	2.3	0.78	95.9	7.0	1.30
Negative glottal stop (n=18; e.g., a niche)	95.0	3.2	1.05	86.7	11.5	2.13
Double cue (n=12) (e.g., grey pin/grape in)	99.2	2.6	0.88	93.6	7.8	1.45
Double cue (Asp−/Glot+) (n=6; e.g., grape in)	100.0	0.0	0.00	99.4	3.0	0.57
Double cue (Asp+/Glot−) (n=6; e.g., grey pin)	98.3	5.3	1.76	87.8	15.1	2.81
All stimuli (n=84)	96.8	2.1	0.70	83.8	5.7	1.06

[a]SD=standard deviation.
[b]SE=standard error.

performance on each stimulus type, is reported in relation to that of the AE group first, then the influence of the presence or absence of the acoustic cues (denoted as positive and negative) on the listeners' performance is reported. Furthermore, Japanese listeners' performance on the subcategories of each stimulus type with positive and negative distinctions is compared. The second part compares the most difficult items for Japanese listeners with those for AE listeners and also examines the performance on the items with irregular acoustic properties (e.g., the positive glottal stop items that do not contain glottal stop) by each language group. Furthermore, whether or not the confidence ratings showed close relations with the error rates of test items was investigated. In the third part, the correlation of the Japanese subjects' performance with their LOR in English-speaking countries is reported.

A. Comparisons of Japanese vs AE groups

1. Overall results and results on each stimulus type

The mean percentage of correct responses by Japanese and AE subjects for each stimulus type is presented in Table III. Nonparametric Median tests evaluated by chi-square were chosen for statistical analysis of the comparisons between the AE and Japanese groups because (a) the numbers of subjects in the two groups were different, and (b) the American group's performance showed ceiling effects. The AE group's performance was significantly higher overall than that of the Japanese group (χ^2 [1df]=12.06, $p<0.001$). Three separate Median tests were conducted to compare the performance on aspiration stimuli (χ^2 [1df]=12.06, $p<0.001$), on glottal stop stimuli (χ^2 [1df]=8.888, $p<0.01$), and on double cue stimuli (χ^2 [1df]=5.004, $p<0.05$). The AE group's performance was significantly better than that of the Japanese group for all three types of contrasts.

The data also show that the Japanese group was much less accurate on Aspiration stimuli (73.1%) than on Glottal stop stimuli (91.3%) or on Double cue stimuli (93.6%) while the AE group performed above 95% on all stimulus types. In addition, one can assume, from the Japanese group's rather high standard deviations (5.7–9.4) compared to those of the AE group (1.6–4.4), that Japanese subjects' performance was more variable than that of the AE group.

Since the AE group's performance was at or near ceiling on all stimulus types, within-group comparisons were performed only for the Japanese data. A nonparametric Friedman's test and Wilcoxon signed-ranks tests were chosen for examining differences in performance across stimulus types because of the unequal variances, skewed distributions, and ceiling effects seen in performance on Glottal stop and on Double cue stimuli. Friedman's test showed that Japanese listeners' performance differed significantly by stimulus types (χ^2 [2df]=46.41, $p<0.001$). Wilcoxon tests further revealed that their performance on Aspiration stimuli was significantly poorer than that on Glottal stop stimuli (z=4.62, $p<0.001$) and on Double cue stimuli (z=4.78, $p<0.01$), while the difference in performance on Glottal stop and Double cue stimuli did not reach significance (z=1.71, $p>0.05$).

2. Results on positive and negative stimuli

Table III also shows the mean percent correct responses by Japanese and AE subjects for Positive and Negative stimuli of each stimulus type. Since the AE group's performance was near ceiling on both Positive and Negative stimuli (though it appears to show slightly better perfor-

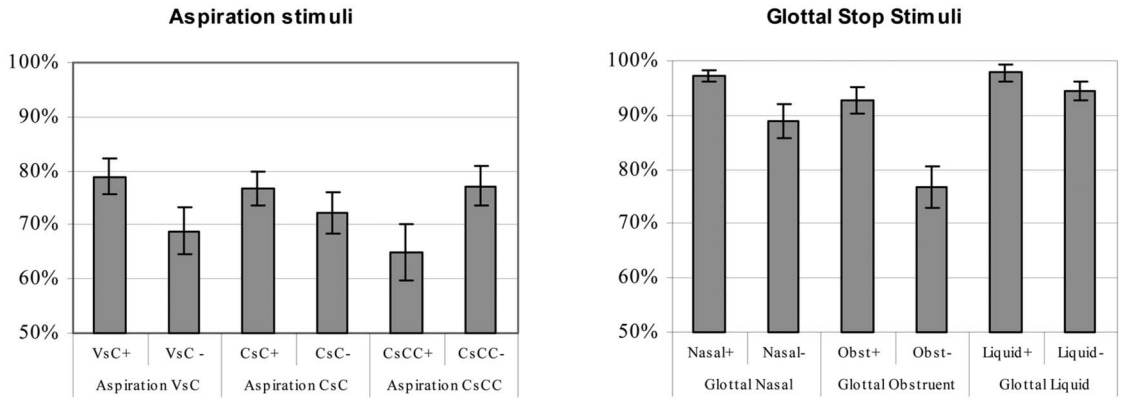


FIG. 2. Japanese listeners' mean percent correct: Aspiration stimuli (left panel) and Glottal stop stimuli (right panel). (Error bar = standard error.)

mance on Positive stimuli than Negative stimuli), statistical comparison between the performance on Positive and Negative stimuli was performed only for the Japanese data.

Wilcoxon signed-ranks tests revealed that Japanese listeners' performance on Positive and Negative stimuli did not differ significantly for Aspiration stimuli ($z=0.55$, $p>0.05$) whereas their performance did differ significantly for Glottal stop stimuli ($z=3.07$, $p<0.01$) and for Double cue stimuli ($z=3.30$, $p<0.01$). Regarding the perception of Glottal stop stimuli, Japanese listeners performed significantly better when the glottal stop cue was present than when it was absent. Their performance on Double cue stimuli also indicated that they performed better on the stimuli with a positive glottal stop cue without aspiration cue than on the stimuli with an aspiration cue but no glottal stop.

Figure 2 shows Japanese listeners' mean percentage of correct responses on subcategorized groups of Aspiration stimuli and Glottal stop stimuli. Here, again, the data for the subcategories of Glottal stop stimuli indicate that the negative stimuli of each subcategory, indicated by "−" in the graph, were always lower than the corresponding positive stimuli, indicated by "+" while those of Aspiration stimuli do not show obvious patterns of such tendencies.[7] Thus, Japanese listeners apparently did not utilize the presence/absence of aspiration cues effectively for word segmentation whereas they counted heavily on the presence of glottal stop cues as word juncture cues.

B. Item analysis

This section examines whether or not Japanese listeners and AE listeners made frequent errors on the same items and whether or not their confidence ratings reflected their difficulty in correctly perceiving the items. Figure 3 illustrates the relations between the error rate of each item and corresponding mean confidence rating by Japanese listeners (left panel) and by AE listeners (right panel) in scatter plots.

The items that were mistaken most by Japanese listeners were *keeps canning* (73.3% errors), *tops crawled* (60.0% errors), *keep stalking* (56.7% errors), and *Lou skis* (56.7% errors), and the items that were most error-inducing for AE listeners were *tea mat* (40% errors), *cook screams* (30% errors), *say vamps* (30% errors), and *cook sprints* (30% errors), showing no overlap of the most difficult items for the two groups. Furthermore, the error-prone items for Japanese listeners were all Aspiration stimuli whereas the error-prone items for AE listeners included both Glottal stop stimuli (*tea mat* and *say vamps*) and Aspiration stimuli (*cook screams* and *cook sprints*).

Although the error rates were not as high as for the items mentioned above, the Japanese group showed higher error rates on some of the items with irregular acoustic properties than on the other items in the same stimulus type. The AE group did not show such a tendency. The error rates of *an ice man* and *save amps*, which were categorized as Positive Glottal stop stimuli but did not have glottal stop/creaky voice, were 13.3% and 26.7%, respectively, for the Japanese group while the error rate for the rest of Positive Glottal stop stimuli was 1.7% on average. On the other hand, the AE group's error rate for *an ice man* was 0% and that for *save amps* was 10% (i.e., only one person answered incorrectly), showing no distinct difference from the mean error rate of the rest of Positive Glottal stop stimuli (0.6%). These results indicate the possibility that both Japanese and AE listeners were using acoustic cues other than glottal stop/creaky voice, although Japanese listeners were more disadvantaged by the lack of the glottal stop/creaky voice cue. For another irregular item *light old*, which was categorized into Double cue (Asp−/Glot+) stimuli but the glottal stop was not distinguishable from the stop silence, the error rate was 3.3% for the Japanese group and 0% for the AE group, showing no greater difficulty than the rest of Asp−/Glot+ stimuli by either the Japanese (6.1%) or by the AE (0%) group.

Spearman rank-order correlations were calculated to examine the relations between the confidence ratings and error rates for both Japanese and AE listeners.[8] The *rho* values were −0.778 for Japanese listeners and −0.377 for AE listen-

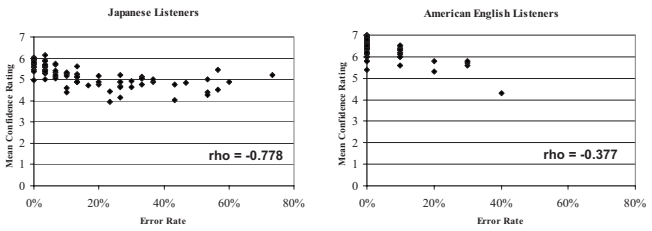


FIG. 3. Mean confidence rating and error rate by Japanese listeners (left panel) and by AE listeners (right panel).

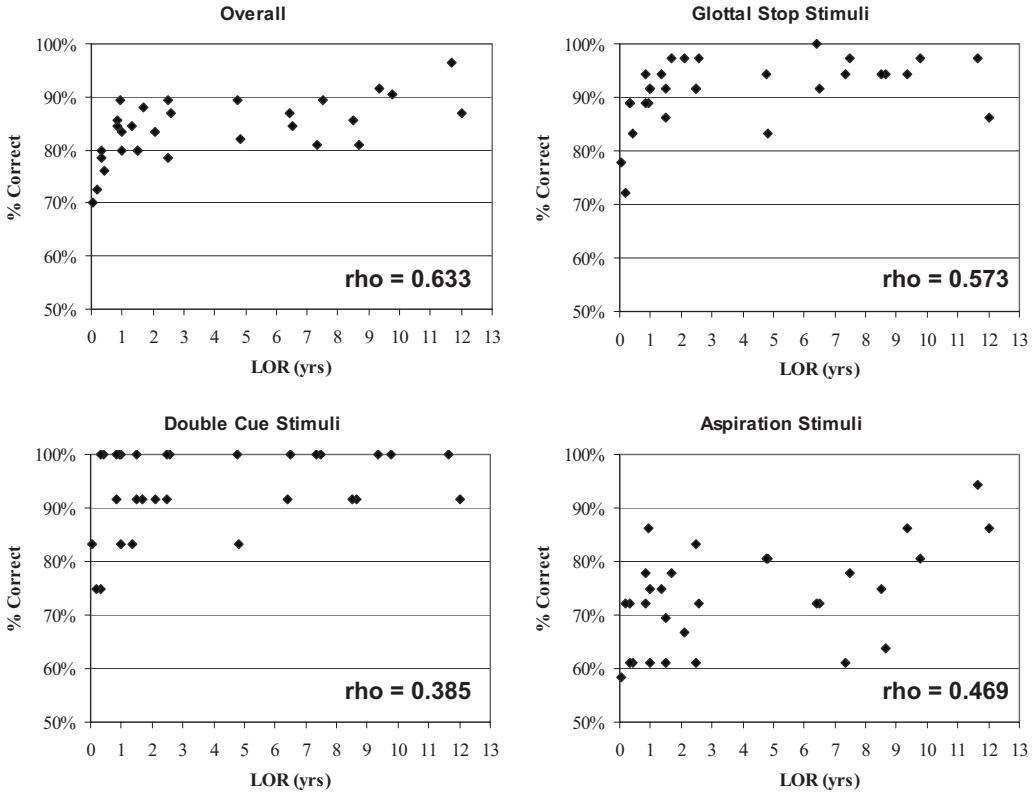


FIG. 4. Japanese listeners' LOR and performance: overall performance (top left panel), performance on Glottal stop stimuli (top right panel), on Double cue stimuli (bottom left panel), and on Aspiration stimuli (bottom right panel).

ers after the correction for tied ranks (Siegel and Castellan, 1988, p. 239). For the AE group, the correlation was limited by the large number of errorless scores (ceiling effects). For the Japanese group, the correlation was quite high, but the range in confidence ratings was limited, with relatively large variations across subjects in ratings for both easy and difficult items. The mean confidence ratings remained quite high even when more than half of the listeners answered incorrectly (see left panel of Fig. 3).

C. Correlation between Japanese listeners' LOR and performance

In order to investigate whether greater L2 experience was associated with listeners' better use of L2 acoustic-phonetic cues for word segmentation, Spearman rank-order correlations between Japanese listeners' LOR in English-speaking countries and their accuracy in the experiment were computed. Figure 4 illustrates the relation between their LOR and overall performance and those between LOR and performance on each of the three different stimulus types. All correlations were positive, although some were rather low, due to ceiling effects that restricted variability. Overall accuracy rates (top left panel) (rho=+0.633) increased rapidly over the first couple of years but flattened after that without reaching ceiling. However, performance seems to have increased again for participants with LOR of 9 years or greater. Performance on Glottal stop stimuli (top right panel) shows a similar tendency to that for overall performance (*rho*=+0.573), demonstrating a rapid increase in response accuracy in the first couple of years of LOR with rates near the ceiling after that. The correlations of LOR with performance on Double cue stimuli (bottom left panel) (*rho*=+0.385) and Aspiration stimuli (bottom right panel) (*rho*=+0.469) were lower than for overall performance and Glottal stop stimuli. Performance on Double cue stimuli showed ceiling effects because many subjects obtained perfect scores regardless of their LORs, although subjects with lower LOR varied more among themselves. For Aspiration stimuli, accuracy rates were widely distributed across LOR, although the subjects with more than 9 years of LOR tended to have higher scores than the rest of the subjects. Comparing these results with overall performance, it appears that the rapid overall improvement in the first few years of LOR is attributable to L2 listeners picking up on the glottal stop cue while the late improvement (i.e., the higher performance after 9 years of LOR) is due to their acquiring the ability to tap into the aspiration cue.

D. Effects of familiarity with phrases on Japanese listeners' performance

The results for the familiarity test for Aspiration stimulus pairs and Glottal stop stimulus pairs by Japanese listeners are summarized in Table IV. As can be seen, the numbers of more familiar items for Positive stimuli and for Negative stimuli in both Aspiration and Glottal stop stimuli were almost evenly distributed, showing no strong bias toward either Positive or Negative stimuli.

For Aspiration stimulus pairs, a tendency that may indicate response bias favoring more familiar items in perfor-

TABLE IV. Results of Japanese listeners' familiarity questionnaire.

Stimulus pair type	Familiarity answer	No. of pairs	No. of more familiar item better identified	Example (% correct in parentheses)
Aspiration stimulus pairs (18 pairs)	Positive item as more familiar	8	7/8 (87.5%)	*loose keys* vs *Lou skis* (90.0%) vs (43.3%)
	Negative item as more familiar	10	7/10 (70%)	*keep scanning* vs *keeps canning* (76.7%) vs (26.7%)
Glottal stop stimulus pairs (18 pairs)	Positive item as more familiar	10	9/10 (90%)	*grave at* vs *grey vat* (100%) vs (66.7%)
	Negative item as more familiar	8	0/8 (0%)	*we vend* vs *weave end* (80.0%) vs (93.3%)
Double cue stimulus pairs (6 pairs)	Asp−/Glot+ item as more familiar	3	2/3 (66.7%) (1 pair was both 100% correct)	*make art* vs *may cart* (100%) vs (73.3%)
	Asp+/Glot− item as more familiar	3	0/3 (0%)	*grey pin* vs *grape in* (73.3%) vs (100%)

mance was observed. The more familiar items tended to be better identified than their counterparts. Seven out of eight positive items and seven out of ten negative items that were chosen as more familiar had higher accuracy rates than the corresponding members of the pairs did. This means that Japanese listeners chose the more familiar items as the answers in the experiment significantly more often than their less familiar counterparts (14/18, significantly different from chance [$p<0.02$] by the binomial expansion). This tendency was not seen in Glottal stop stimulus pairs. In Glottal stop pairs, none of the more familiar negative items was identified more accurately than their positive counterparts while nine out of ten more familiar positive items were better identified than their negative counterparts. The better performance on Positive Glottal stop stimuli than on their negative counterparts, regardless of familiarity, indicates a lack of response bias for familiar items in identification of word onsets cued by glottal stops (9/18, not significantly different from chance). A similar pattern is seen in Double cue stimulus pairs. None of Asp+/Glot− items were better identified than their Asp−/Glot+ counterparts irrespective of their relative familiarity with their counterparts (2/6, not significantly different from chance). These results suggest that familiarity with words or phrases does not seriously affect how listeners segment speech when they can effectively use allophonic segmentation cues, but may affect segmentation when they cannot take advantage of the allophonic cues.

IV. DISCUSSION

This section summarizes the results of the study, offering answers to the three research questions presented in the Introduction, and then considers the limitations of the study and some directions for future studies.

A. Perceptual differences between Japanese and AE listeners

The results clearly indicated that Japanese L2 listeners, as a group, perceived English word boundaries more poorly than native AE listeners in many ways. First, Japanese listeners' overall performance was significantly worse than that of AE listeners, suggesting that Japanese listeners could not use the available cues as effectively as native listeners did. Second, Japanese listeners showed a significant difference in their performance on different stimulus types while AE listeners' performance was near ceiling on all stimulus types. Japanese performance on Aspiration stimuli was significantly worse than that on Glottal stop stimuli and on Double cue stimuli while the AE group's correct rates were consistently above 95%. Third, the items that were mistaken most by Japanese listeners and by AE listeners did not overlap. The four most error-prone items for Japanese listeners were all Aspiration stimuli, whereas those of AE listeners included both Glottal stop stimuli and Aspiration stimuli, implying different perceptual patterns by Japanese and AE listeners. Fourth, Japanese listeners seemed to have a poorer sense than AE listeners of how well they were able to perceive the word boundaries. The scatter plots of confidence ratings and correct performance showed that the Japanese listeners' confidence ratings did not decline much even when more than half of the listeners answered incorrectly, while the AE listeners' confidence ratings better reflected their error rates on the few items that were difficult. Finally, Japanese listeners seemed to be more disadvantaged than native AE listeners by the lack of the most distinctive acoustic cue for segmentation. Japanese listeners misperceived the word boundaries of the irregular Positive Glottal stop stimuli that did not have glottal stop/creaky voice more often than the stimuli with glottal stop/creaky voice present, while AE listeners did not show such a difference. These results suggest that when a reliable acoustic cue is absent, Japanese listeners are not capable of taking advantage of the remaining acoustic cues to the same extent as AE listeners. These results generally replicate those reported by Altenberg (2005) for Spanish late L2 learners of English on the same materials spoken by the different speaker.

B. Relative perceptual problems with different types of segmentation cues

Japanese listeners made significantly more errors on pairs differentiated by aspiration cues than they did on pairs differentiated by glottal stop cues. Furthermore, the data comparing Japanese listeners' performance on Positive and Negative stimuli within each stimulus type show that they performed significantly better when the glottal stop cues were present than when they were absent. However, the presence or absence of the aspiration cues did not significantly affect their performance. This pattern of results replicates those reported by Altenberg (2005) for Spanish L2 learners with the exception that her subjects showed better relative performance on both Positive Aspiration stimuli and Positive Glottal stop stimuli.

We can tentatively conclude that, independent of differences in the use of aspiration and glottal stops in the L1, aspiration cues (VOT and correlated duration parameters) are harder to learn by L2 learners than are glottal stop cues (and related parameters) for segmentation. In the course of discussing the cause of relative difficulty for Spanish listeners, Altenberg (2005) explored the concept of markedness as one of the possible factors. That is, "glottal stop insertion is less marked and hence easier to acquire than aspiration" (pp. 344–345). It has been observed by speech scientists that initiating phonation with a glottal stop (Borden *et al.*, 2003) and inserting a glottal stop at the boundary of abutting words ending and beginning with vowels (Kent and Read, 2002) are quite common by some speakers. Furthermore, glottal stop insertion at a word juncture preceding a word-initial vowel is commonly observed in many languages other than English, Japanese, and Spanish (e.g., Kohler, 1994; Rubach, 2000). Thus, it might have been the case that the less marked glottal insertion was learned more easily as a segmentation cue by Japanese listeners than the aspiration cue.

Another possible explanation may be related to the influence of the L1, or L1 transfer, as also suggested by Altenberg (2005). In comparing the pattern of relative difficulty across Japanese and Spanish listeners the two groups showed comparable performance on Glottal stop and Double cue stimuli, whereas Japanese listeners performed relatively better than Spanish listeners did on Aspiration stimuli (although worse than AE controls). This pattern of relative difficulty in utilizing aspiration cues corresponds to the extent of word-initial aspiration of voiceless stops in the three languages (AE > Japanese > Spanish), as discussed earlier. It has to be noted, however, that the data reported here and those of Altenberg (2005) are not completely suitable for comparison because of differences with regard to subject inclusion criteria. Altenberg (2005) tested college-level English as a second language students, whereas the present study included participants who ranged from those with almost no immersion experience to those with more than 10 years of immersion experience.

It is also useful to consider if constraints of L1 syllable structure affect Japanese listeners' L2 perception of word boundaries. Japanese syllable structure does not allow a consonant cluster within a syllable because the majority of Japanese syllables are CV syllables, with some exceptions of CVC syllables involving syllable-final moraic consonants (moraic final nasal and geminate consonants). If L1 syllable constraints influence the perception of L2 syllable patterns, therefore, one may predict that Japanese listeners would have more difficulty in segmenting words correctly as the consonant cluster at the word juncture in Aspiration stimuli became more complex. The results, however, did not show any obvious tendency. In the Aspiration stimuli, the consonant cluster at word juncture becomes more complex in the following order: VsC group, CsC group, and CsCC group. The mean accuracies on the above groups by Japanese listeners were 73.9%, 74.4%, and 71.1%, respectively, and did not differ from each other.[9]

Thus, the results regarding L1 transfer seem to be mixed. The relation between the degree of aspiration in L1 and that in L2 seems to suggest L1 transfer in perceiving L2 word boundaries whereas L1 syllable patterns do not seem to affect the perception of L2 syllable at word boundaries.

C. Relationship between immersion experience and use of allophonic cues

The positive relationships between Japanese listeners' performance and their LOR suggest that increased English experience does improve Japanese listeners' use of English allophonic cues. Their overall response accuracy increased as a function of their LOR, although it was not linearly related. A rapid difference in response accuracy across the first 1–3 years of LOR followed by a plateau was observed. Performance appeared to improve again after 9–10 years, although the sampling of participants with long LOR was limited. While Japanese listeners seem to learn to exploit glottal stop cues for English word segmentation within the first few years of residence, their mastery of aspiration cues may come much later. The pattern seen in the overall performance seems to reflect this "time lag" of learning different allophonic cues. These tendencies need to be followed up in future studies with larger numbers of participants with extensive immersion experience.

D. Other issues: Limitations and future studies

In addition to the findings related to the research questions, the present study suggested that familiarity with words or phrases may have affected Japanese listeners' responses in the experiment when they were not using the allophonic cues effectively. The results of the familiarity questionnaire revealed that for Aspiration stimuli, on which Japanese listeners performed significantly worse than the other types of stimuli, items that were chosen by Japanese listeners as more familiar had higher accuracy rates than the corresponding members of the pairs, indicating a response bias toward the familiar items. Since this tendency was not seen in Glottal stop stimuli, the results imply that familiarity with words or phrases affected performance only when Japanese listeners

could not fully use the acoustic-phonetic segmentation cues. This agrees with the finding reported by Mattys and Melhorn (2007) that lexical cues influence segmentation decisions when allophonic cues are ambiguous. Altenberg (2005) attempted to minimize lexical effects when constructing the stimuli; however, the present results suggest that it may not be possible to eliminate the effects of lexical bias completely from word segmentation studies of this kind. Future studies must take these effects into consideration when interpreting results.

As mentioned in the Introduction, cross-language research on L2 word segmentation is not extensive. There are fewer studies dealing with individual acoustic-phonetic cues than those examining rhythmic cues. The present study examined the use of only two of the many allophonic cues used in English word segmentation. Furthermore, the issue of coarticulatory effects across word boundaries, such as assimilation and deletion, which may cause significant changes in acoustic/phonetic properties of phonetic segments, has not been addressed sufficiently in L2 word segmentation studies, in spite of the fact that such coarticulatory variations commonly occur in connected speech. Future studies should examine these issues. In the process of L2 learning, non-native listeners must learn to perceive the distinctive segments (phoneme contrasts), the phonotactic patterns, and the allophonic variations of phonetic segments that serve as segmentation cues. Research on all aspects of L2 phonological learning will provide us with a better view of the problems associated with late language learning and guide our efforts to improve L2 language processing through intensive L2 training methods.

ACKNOWLEDGMENTS

We thank Evelyn Altenberg for her helpful comments and suggestions in preparing this manuscript. This research was supported in part by a grant to W.S. (NIH Grant No. DC00323).

APPENDIX: STIMULUS PAIRS USED IN THE EXPERIMENT [TAKEN FROM ALTENBERG (2005)]

ASPIRATION STIMULUS PAIRS

Positive stimuli	**Negative stimuli**
VsC group	
loose pills	Lou spills
loose tops	Lou stops
loose keys	Lou skis
lace peach	lay speech
lace table	lay stable
lace car	lay scar
CsC group	
keeps parking	keep sparking
keeps talking	keep stalking
keeps canning	keep scanning
chief's port	chief sport
chief's tar	chief star
chief's cool	chief school
CsCC group	
cook's prints	cook sprints
cook's truck	cook struck
cook's creams	cook screams
tops pry	top spry
tops trains	top strains
tops crawled	top scrawled

GLOTTAL STOP STIMULUS PAIRS

Positive stimuli	**Negative stimuli**
Nasal group	
seen either	see neither
an itch	a niche
an ice man	a nice man
seem able	see Mabel
team at	tea mat
claim annual	clay manual
Obstruent group	
wife ill	why fill
loaf ate	low fate
loaf ailing	low failing
grave at	grey vat
weave end	we vend
save amps	say vamps
Liquid group	
I'll earn	I learn
sail east	say least
seal of	see love
beer old	be rolled
cure ache	cue rake
beer ice	be rice

DOUBLE CUE STIMULUS PAIRS

Asp−/Glot+ Stimuli	**Asp+/Glot− Stimuli**
grape in	grey pin
weep at	we pat
light old	lie told
might owe	my toe
make art	may cart
like old	lie cold

Stimulus pairs used in the familiarization task [taken from Nakatani and Dukes (1977) except for *ice cream/I scream*]

ice cream/I scream	lawn chair/launch air
we'll own/we loan	stayed ill/stay dill
plate ought/play taught	no notion/known ocean

[1]Throughout this study, we will refer to these cues as aspiration cues and glottal stop cues, although an acoustic analysis of the stimuli revealed that there were other acoustic parameters that differentiated word-initial consonants and vowels.

[2]Although Altenberg (2005) selected the stimuli to minimize lexical influences on the experiment by using native speakers' naturalness rating, it cannot be ascertained whether L2 listeners' lexical bias contributed to the results. Altenberg (2005) reported that word frequency was not likely to affect the results by comparing the average frequency counts of words in

each stimulus type and the corresponding Spanish listeners' performance (pp. 348–349). In the present study, however, the familiarity test was designed for a more direct investigation of the possible correlation between participants' familiarity with words or phrases and their performance on the segmentation task.

[3]All Japanese participants were asked to report experience in staying in English-speaking countries only if it lasted for 1 month or longer, except for one participant who had come to the United States only 2 weeks before the test.

[4]Although it can also be pronounced as /niʃ/, the word *niche* was pronounced as /nɪtʃ/ in this experiment in order to make the phrase *a niche* nearly homophonous to the phrase *an itch*.

[5]The data for the Liquid group were not included in the statistic analysis because of the non-segmentability of the liquid portion in Positive stimuli.

[6]The original Altenberg experiment (2005) tested the participants in small groups in a quiet room with a fixed interval in which to respond on answer sheets.

[7]Wilcoxon signed-ranks tests showed that Japanese listeners' performance difference between Positive and Negative stimuli did not reach significance in any of subcategorized groups of Aspiration stimuli ($z=1.76$ for VsC group; $z=0.94$ for CsC group; $z=1.43$ for CsCC group).

[8]Because of the nonlinearity of the relationships, the non-parametric Spearman test that does not assume normal distribution or linearlity of the relationship was employed (Siegel and Castellan, 1988).

[9]Wilcoxon signed-ranks tests revealed that none of the Japanese listeners' performance differences between subcategorized groups of aspiration stimuli reached significance ($z=0.08$ for VsC vs CsC group; $z=1.09$ for VsC vs CsCC group; $z=1.13$ for CsC vs CsCC group).

Altenberg, E. P. (**2005**). "The perception of word boundaries in a second language," Second Lang. Res. **21**, 325–358.

Bloch, B. (**1950**). "Studies in colloquial Japanese IV: Phonemics," Language **26**, 86–125.

Borden, G. J., Harris, K. S., and Raphael, L. J. (**2003**). *Speech Science Primer*, 4th ed. (Lippincott Williams and Wilkins, Philadelphia, PA).

Bradley, D. C., Sanchez-Casas, R. M., and Garcia-Albea, J. E. (**1993**). "The status of the syllable in the perception of Spanish and English," Lang. Cognit. Processes **8**, 197–233.

Christie, W. M. (**1974**). "Some cues for syllable juncture perception in English," J. Acoust. Soc. Am. **55**, 819–821.

Church, K. (**1987**). "Phonological parsing and lexical retrieval," Cognition **25**, 53–69.

Cutler, A., and Butterfield, S. (**1992**). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," J. Mem. Lang. **31**, 218–236.

Cutler, A., and Norris, D. (**1988**). "The role of strong syllables in segmentation for lexical access," J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121.

Cutler, A., and Otake, T. (**1994**). "Mora or phoneme? Further evidence for language-specific listening," J. Mem. Lang. **33**, 824–844.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (**1986**). "The syllables' differing role in the segmentation of French and English," J. Mem. Lang. **25**, 385–400.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (**1992**). "Monolingual nature of speech segmentation by bilinguals," Cogn. Psychol. **24**, 381–410.

Fougeron, C., and Keating, P. A. (**1997**). "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. **101**, 3728–3740.

Goetry, V., and Kolinsky, R. (**2000**). "The role of rhythmic cues for speech segmentation in monolingual and bilingual listeners," Psychol. Belg. **40**, 115–152.

Gow, D. W., and Gordon, P. C. (**1995**). "Lexical and prelexical influences on word segmentation: Evidence from priming," J. Exp. Psychol. Hum. Percept. Perform. **21**, 344–359.

Hattori, S. (**1951**). *Onseigaku (Phonetics)* (Iwanami, Tokyo).

Homma, Y. (**1980**). "Voice onset time in Japanese stops," Onsei Gakkai Kaiho **163**, 7–9.

Kawakami, S. (**1977**). *Nihongo Onsei Gaisetsu (An Outline of Japanese Phonetics)* (Ofusha, Tokyo).

Kent, R., and Read, C. (**2002**). *The Acoustic Analysis of Speech*, 2nd ed. (Singular, San Diego, CA).

Kohler, K. J. (**1994**). "Glottal stops and glottalization in German," Phonetica **51**, 38–51.

Mattys, S., and Melhorn, J. F. (**2007**). "Sentential, lexical, and acoustic effects on the perception of word boundaries," J. Acoust. Soc. Am. **122**, 554–567.

Mattys, S., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework," J. Exp. Psychol. Gen. **134**, 477–500.

McQueen, J. M. (**1998**). "Segmentation of continuous speech using phonotactics," J. Mem. Lang. **39**, 21–46.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., and Segui, J. (**1981**). "The syllable's role in speech segmentation," J. Verbal Learn. Verbal Behav. **20**, 298–305.

Nakatani, L. H., and Dukes, K. D. (**1977**). "Locus of segmental cues for word juncture," J. Acoust. Soc. Am. **62**, 714–719.

Oller, D. K. (**1973**). "The effect of position in utterance on speech segment duration in English," J. Acoust. Soc. Am. **54**, 1235–1247.

Otake, T., Hatano, G., Cutler, A., and Mehler, J. (**1993**). "Mora or syllable? Speech segmentation in Japanese," J. Mem. Lang. **32**, 258–278.

Riney, T. J., Takagi, N., Ota, K., and Uchida, Y. (**2007**). "The intermediate degree of VOT in Japanese initial voiceless stops," J. Phonetics **35**, 439–443.

Rubach, J. (**2000**). "Glide and glottal stop insertion in Slavic languages: A DOT analysis," Linguistic Inquiry **31**, 271–317.

Sanders, L. D., and Neville, H. J. (**2000**). "Lexical, syntactic, and stress-pattern cues for speech segmentation," J. Speech Lang. Hear. Res. **43**, 1302–1321.

Sanders, L. D., and Neville, H. J. (**2003a**). "An ERP study of continuous speech processing II. Segmentation, semantics, and syntax in non-native speakers," Brain Res. Cognit. Brain Res. **15**, 214–227.

Sanders, L. D., and Neville, H. J. (**2003b**). "An ERP study of continuous speech processing I. Segmentation, semantics, and syntax in native speakers," Brain Res. Cognit. Brain Res. **15**, 228–240.

Sanders, L. D., Neville, H. J., and Woldorff, M. G. (**2002**). "Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern," J. Speech Lang. Hear. Res. **45**, 519–530.

Siegel, S., and Castellan, N. J., Jr. (**1988**). *Nonparametric Statistics for the Behavioral Sciences*, 2nd ed. (McGraw-Hill, Boston, MA).

Stockwell, R. P., Bowen, J. D., and Silva-Fuenzalida, I. (**1956**). "Spanish juncture and intonation," Language **32**, 641–665.

Vance, T. J. (**1987**). *An introduction to Japanese Phonology* (State University of New York Press, New York).

Williams, L. (**1977**). "The voicing contrast in Spanish," J. Phonetics **5**, 169–184.

Processing of fast speech by elderly listeners

Esther Janse[a)]
Utrecht Institute of Linguistics OTS and Max Planck Institute for Psycholinguistics Nijmegen, Janskerkhof 13, 3512 BL Utrecht, The Netherlands

(Received 8 April 2008; revised 21 January 2009; accepted 25 January 2009)

This study investigates the relative contributions of auditory and cognitive factors to the common finding that an increase in speech rate affects elderly listeners more than young listeners. Since a direct relation between non-auditory factors, such as age-related cognitive slowing, and fast speech performance has been difficult to demonstrate, the present study took an on-line, rather than off-line, approach and focused on processing time. Elderly and young listeners were presented with speech at two rates of time compression and were asked to detect pre-assigned target words as quickly as possible. A number of auditory and cognitive measures were entered in a statistical model as predictors of elderly participants' fast speech performance: hearing acuity, an information processing rate measure, and two measures of reading speed. The results showed that hearing loss played a primary role in explaining elderly listeners' increased difficulty with fast speech. However, non-auditory factors such as reading speed and the extent to which participants were affected by increased rate of presentation in a visual analog of the listening experiment also predicted fast speech performance differences among the elderly participants. These on-line results confirm that slowed information processing is indeed part of elderly listeners' problem keeping up with fast language. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3082117]

PACS number(s): 43.71.Lz [MSS] Pages: 2361–2373

I. INTRODUCTION

When people age, listening to speech can become an effortful task. Frequent complaints of elderly listeners are that they have trouble listening to speech against a background of noise or competing speech, or that their interlocutors speak too fast, or do not articulate clearly enough. The high-frequency peripheral hearing loss typically found in a large sample of elderly people (Willott, 1991) can be held as the primary cause of the speech perception problems. Additional age-related causes are central auditory processing impairments and (general) cognitive decline (information processing speed or attentional problems). The central auditory problems have been reported to be caused by peripheral hearing loss: hearing loss can impair auditory functions as frequency and temporal resolution (Glasberg and Moore, 1986; Sommers and Humes, 1993; George *et al.*, 2007). Note, however, that there is also overlap between the "central auditory" and "cognitive" accounts (Jerger *et al.*, 1991). Either way, the central auditory account seems to be a bridge between whatever is received as auditory input and how it is processed. Humes (2005) noted, on the basis of data from a large group of elderly listeners, that performance on a battery of auditory processing measures was systematically related to individual differences in cognitive, rather than auditory, function. For recognition of time-compressed words, hearing loss accounted for the greatest part of the variance, but there was a significant correlation between (non-verbal) IQ and the ability to perceive time-compressed words as well. Similar relations between auditory processing impairments and higher cognitive-level impairments can be found in the domains of aphasia (Fink *et al.*, 2006) and dyslexia (Tallal, 1980; Wright *et al.*, 1997; Witton *et al.*, 1998, 2002; and Rosen, 2003 for an overview).

One approach to tease apart the relative contributions of age-related changes in hearing and cognitive factors to the increased problems elderly adults may have with speeded speech is to include several groups of listeners (Gordon-Salant and Fitzgibbons, 1993, 1995, 1999, 2001): young listeners with and without hearing loss and elderly listeners with and without hearing loss. In this way, the effects of aging and of hearing loss can be investigated separately. Younger adults with hearing loss also show substantial decrements in recognizing time-compressed speech compared to age-matched listeners with normal hearing. Gordon-Salant and Fitzgibbons (2001), Dubno *et al.* (1987), and Turner *et al.* (1997) attributed these problems to hearing impaired listeners' capacity to use brief acoustic cues. Nevertheless, Gordon-Salant and Fitzgibbons (1993, 2001) also showed that the problems in recognizing time-compressed speech that are attributed to age are independent of those attributed to hearing loss.

One other approach to evaluate the contribution of age-related hearing and cognitive factors to the increased problems elderly have with speeded speech is to speed up some parts of the signal more than others (Gordon-Salant and Fitzgibbons, 2001; Schneider *et al.*, 2005), or to allow listeners to "regain lost time" by introducing silent intervals at strategic points in the time-compressed speech material. In this latter way, listeners get time to catch up and process the information just presented at a high rate (Wingfield *et al.*, 1999). Even though allowing more time to catch up processing helped both elderly and young listeners, the elderly listeners, unlike the young, could not be brought back to base-

[a)]Electronic mail: E.Janse1@uu.nl

line (uncompressed) performance. These results could be taken to suggest that both auditory and cognitive factors contribute to elderly listeners' problems with fast speech. However, Schneider *et al.* (2005) argued for an auditory-only explanation, based on their results showing that the method of time compression matters. They referred to the findings of Gordon-Salant and Fitzgibbons (2001) that older adults find it especially difficult to deal with selective time compression of consonants because identification of consonants may depend on rapid formant transitions. Where most studies have used methods of time compression that involve removing speech segments without regard to their informational content, Schneider *et al.* (2005) used a selective method of time compression mainly affecting pauses or steady-state portions of vowels. This latter method of time compression thus presents the listener with a high rate of information content, while producing minimal acoustic distortion of the speech signal. If speech is speeded in such a selective way, elderly listeners are not affected more by an increase in rate than younger listeners (Schneider *et al.*, 2005). In this view then, an inability of the older adult's auditory system to cope with speed-induced stimulus degradation thus suffices to explain the problem, leaving little or no room for cognitive slowing.

Given the difficulty to pinpoint the role of central auditory processing problems, a third approach has tried to account for the age group × speed issue by investigating whether age-related cognitive slowing problems are modality-specific or not. In order to decide between central auditory and general cognitive explanations for elderly listeners' problems with fast speech, Humes *et al.* (2007) investigated whether cognitive processing is affected more in the auditory modality than in other modalities. Humes *et al.* (2007) compared speeded-speech performance to a visual text-based analog of their auditory test to investigate whether elderly participants were also more affected by increased presentation rate than younger participants when reading sentences. If amodal cognitive decline underlies performance on both reading and listening, performance should be strongly correlated because of this common underlying factor. Correlations across modalities had been found in people's ability to "fill up the gaps:" the results obtained with the text reception threshold test as an analog of the speech reception threshold test showed significant correlations between a participant's ability to identify auditory sentences in noise and written sentences masked by a bar pattern (Zekveld *et al.*, 2007; George *et al.*, 2007). As found by Zekveld *et al.* (2007), Humes *et al.* (2007) found correlations of performance for similar tasks across modalities. Importantly, however, with respect to presentation rate, Humes *et al.* (2007) found that older adults, regardless of hearing status, were more affected by the speeded speech than young adults, whereas there were no differences between groups regarding the effect of rate of visual presentation. This latter result agreed with results by Spehar *et al.* (2004) who found equal effects of speeding for older and younger adults in a test where words had to be identified on the basis of the visual signal only. The absence of a differential rate effect in reading for the elderly in Humes *et al.* (2007) is surprising given the claim that aging slows down the cumulative duration of all processes involved in visual word recognition (ranging from basic visual analysis of letter strings to activation of lexical presentations) with a factor of approximately 1.5 (Myerson *et al.*, 1992). Myerson *et al.* (1992) argued for this slowing factor on the basis of a meta-analysis of (visual) lexical decision studies. Since a similar performance ratio of older and younger adults had been found for naming studies (Balota and Duchek, 1988), Myerson *et al.* (1992) argued for general lexical slowing in aging. In other words, age not only affects sensory decoding of language (either in the auditory or visual domain) but also affects lexical activation patterns. Likewise, at the higher level of sentence comprehension, aging has been shown to delay semantic integration (Federmeier *et al.*, 2002, 2003). Nevertheless, studies on aging and speech perception have often focused on elderly listeners' greater use of sentence context as a means to compensate for their poorer hearing (Nittrouer and Boothroyd, 1990; Pichora-Fuller *et al.*, 1995; Sommers and Danielson, 1999). However, using context to fill up the gaps is a cognitive mechanism requiring mental resources (McCoy *et al.*, 2005). Thus, when we aim to investigate the *relative* contributions of hearing and cognitive factors to elderly adults' problem of keeping up with a fast rate of speech, allowing elderly participants unlimited time to come up with a response may blur our conception of why perceptual processing itself has become more effortful and on why (cognitive) repair strategies are necessary.

Summing up, the literature overview above shows that it has been relatively difficult to establish which factors determine one's ability to keep up with a fast speech rate. The role of (peripheral) auditory decline seems relatively clear: having to decode an impoverished speech signal requires more time and effort, which becomes all the more problematic when speech rate is fast. However, the direct relation between non-auditory factors, such as cognitive slowing, and fast speech performance has been much more difficult to demonstrate. In order to obtain a better understanding of elderly listeners' increased difficulty with time-compressed speech, the present study took an on-line approach. Rather than focusing on the outcome of the recognition process (in terms of recall or correct identification of sentences), the present study focuses on speech processing itself. Investigating just the result of speech processing might obscure which processes led to this result. Particularly if decreased information processing rate is one of the potential underlying causes, it makes sense to tap into speech processing as the speech signal unfolds and to look at processing time, in combination with the result of processing.

Therefore, in the present study, an on-line speech processing approach was chosen to investigate how quickly listeners would be able to detect target words in running speech. Elderly and young listeners were presented with fast speech at two rates of time compression and were asked to detect pre-assigned target words as quickly as possible. An age group × speed interaction was anticipated. The aim of the present study was to establish, within the group of elderly listeners, which factors (and to what extent) determine an individual elderly adult's ability to keep up with a fast speech rate. To this end, elderly listeners were recruited with

varying amounts of (age-induced) hearing loss. It was established how well individual hearing acuity and a number of non-auditory measures each predict elderly participants' word detection performance. The non-auditory measures involved a general measure of information processing speed [the digit symbol substitution (DSS) task] and two measures of the elderly participants' reading speed. Importantly, a number of studies have failed to find a link between elderly listeners' speech performance in taxing conditions and DSS score (Tun, 1998; Sommers and Danielson, 1999), thus challenging the impact of age-related general slowing on speech processing. This failure to find a relation might be due to the fact that most studies have used relatively short sentences (often from the revised sentence perception in noise materials) and have allowed listeners' unlimited response time to reproduce the presented sentence. The present focus on processing time may clarify whether general information processing speed, as measured by DSS performance, predicts fast speech performance in elderly participants.

Individual reading performance, as a measure of information processing rate specific to *language*, was expected to be a better predictor of fast speech performance than the more general DSS measure. If activation of lexical representations is slowed by age (due to general cognitive slowing), this should show up as a differential effect of increased rate of presentation on *reading* performance of the elderly participants as well (an age group × speed interaction in reading). Two complementary measures of individual reading speed were selected to investigate which measure was a better predictor of one's difficulty keeping up with fast speech rates: either self-preferred reading rate (in a self-paced reading experiment) or a measure of the effect of increased rate in a word detection study in which two fast reading rates were imposed on the participants. This latter word detection experiment was a visual analog of the word detection study with fast speech: similar compression rates were used in the two sensory modalities (listening and reading) to investigate whether there would be a shared underlying component in the effect of increased speed. In both reading studies (visual word detection and self-paced reading), presentation of sentences was done in a word-by-word fashion to make reading more like listening with respect to the transient nature of speech. Obviously, word-by-word reading is far from normal or typical reading where readers are able to look back and look ahead. Nevertheless, word-by-word reading resembles speech processing more than a more typical way of reading where all words are present at the same time.

The present study was set up to test the hypothesis that both auditory and non-auditory measures predict elderly listeners' relative difficulty with fast speech rates. Hearing acuity may be the most important one in determining individual difficulty in keeping up with fast rates of speech. However, measures of general cognitive slowing and of reading rate were also expected to be significant predictors of how well one can keep up with fast rates of speech. In other words, this on-line study was expected to show that non-auditory measures indeed contribute to elderly listeners' difficulty in keeping up with fast speech rates.

II. AUDITORY WORD DETECTION STUDY

A. Time compression

In line with many previous studies on fast speech and aging (Tun *et al.*, 1992; Gordon-Salant and Fitzgibbons, 1993, 1995, 1999, 2001; Vaughan and Letowski, 1997; Tun, 1998; Wingfield *et al.*, 1999; Schneider *et al.*, 2005), the present study focuses on artificially time-compressed speech, and not on naturally produced fast speech. The motivation for this choice is that artificial speeding leaves articulation clarity intact, whereas in naturally produced fast speech, higher rates are (almost inevitably) accompanied by greater slurring. Given that naturally produced fast speech was shown to be less intelligible than artificially time-compressed speech (Janse, 2004), and given the literature on "speaking clearly for the hard of hearing" (Picheny *et al.*, 1985, 1986, 1989; Uchanski *et al.*, 1996), artificially speeded speech was used to avoid further complication from slurred or unclear speech.

As mentioned in the Introduction, Schneider *et al.* (2005) attributed elderly listeners' problems with time-compressed speech mainly to their inability to cope with the acoustic artifacts of the time-compression technique used in most studies. They concluded this on the basis of a comparison between the sampling technique (in which short, e.g., 10-ms, segments are deleted periodically) and a more refined technique in which only steady-state portions of the signal were removed (i.e., pauses or gaps between words or portions of steady-state vowels). However, if no age differences are found with the more refined technique of time compression, this could be due either to the absence of acoustic artifacts in, e.g., consonant transitions (as argued by Schneider *et al.*, 2005) or by the non-linear (and thus less severe) way of speeding the sentence. Compressing pauses more than the remaining speech is a form of non-linear time compression that will always lead to better intelligibility than can be obtained with linear compression because the speech itself is less compressed (Janse, 2003, and compare the results obtained with time-compression algorithm Mach 1; Covell *et al.*, 1998). The results of the study of Schneider *et al.* (2005) do not rule out the possibility that an age × speed interaction may be found in terms of perceptual effort, or at more taxing time-compression rates.

The present study compares two methods of time compression: the sampling method as a rather crude method of deleting speech material that involves audible artifacts and the pitch-synchronous-overlap-add (PSOLA) method. PSOLA (Moulines and Charpentier, 1990), because of its high-quality time-scale manipulation, has become the technique incorporated in most speech editing programs (COOLEDIT, PRAAT, and AUDACITY). First, a pitch detection algorithm places markers at consecutive pitch periods (also in the unvoiced portions). The signal is decomposed into separate but overlapping (Hanning) windows, with window maxima at each pitch mark location. Window length is usually twice the length of one pitch period (yielding overlap in consecutive windows). Time compression is accomplished by deleting pitch periods. The crucial asset of PSOLA time compression is that in constructing the new waveform, the

speech signal of one descending window ramp is added to that of the next ascending window ramp (with, e.g., one intervening window deleted). This means that the information contained in the deleted pitch period is averaged across now overlapping windows, thus avoiding discontinuities and spectral jumps (cf. also Letowski and Poch, 1996 and Vaughan and Letowski, 1997) and thus avoiding nasty acoustic artifacts. Any acoustic distortion can be anticipated to be more detrimental the poorer the hearing level of the listener. Inclusion of these two methods of time compression provides a comparison of their perceptual consequences for elderly listeners. Importantly, time compression was applied linearly with both time-compression techniques: the two techniques differed in whether or not they involved audible discontinuities.

Two rates of time compression were chosen: 1.5 times the original rate (compression to 67%) and 2 times the original rate (compression to 50%). These two rates were chosen on the basis of the results of a pilot listening experiment with elderly listeners. Since this pilot group of listeners, having a mean age of 80 years, hardly thought the experimental task challenging when the materials were presented at 1.5 times the normal rate, and since mean age was expected to be lower for the elderly participants in the present experiment, the 1.5 times normal-rate condition formed the baseline condition against which the fastest rate (2 times normal rate) could be compared.

Time compression was achieved by applying either PSOLA or sampling compression to the original speech materials in the sound-editing software package PRAAT. The sampling version of the 1.5 times rate condition (67%) was made by deleting every third 10-s fragment. The sampling version of the 2 times normal-rate condition (50%) was achieved by deleting every other 10-ms fragment. The sampling method yielded an audibly discontinuous sound, due to waveform discontinuities and spectral jumps at points at which now adjacent fragments were joined.

B. Participants

The elderly participants were recruited via *Hoger Onderwijs Voor Ouderen* ("Higher Education for the Elderly"): an organization linked to several universities in the Netherlands that provides academic courses (on a variety of topics) for people of over 55 years of age. People who had subscribed for a course received an information letter in which they were asked to participate in the study. Given the academic level of the courses, the elderly group was a good match to the young (student) group in terms of education level.

There were 40 elderly participants, aged between 65 and 84, with a mean age of 70 (SE=0.7). Of the elderly participants, 18 held a university degree and 12 had finished a college of higher education (polytechnic). The 20 young participants were all students of Utrecht University (aged between 17 and 26, with a mean age of 21). All participants were native speakers of Dutch and they all received a small payment (students) or present (elderly) for their participation.

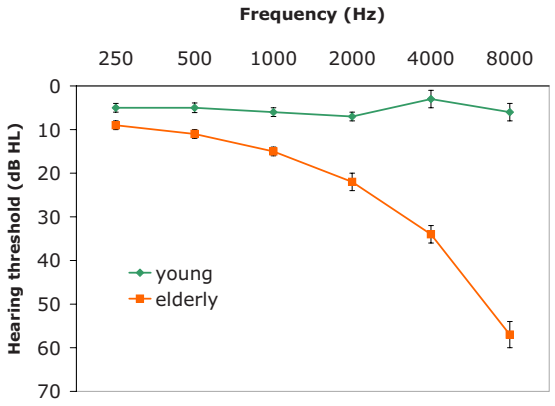


FIG. 1. (Color online) Mean pure-tone thresholds (better ear) for both age groups. Error bars reflect one standard error.

Some of the elderly participants had a hearing aid (only 4 out of 40). They were asked to remove it when participating in the listening experiment. Hearing levels of all participants were measured with a portable Maico ST20 audiometer in a silent booth of the Utrecht Institute of Linguistics Phonetics Laboratory. Figure 1 displays the mean threshold values (in decibels) measured in the better ear for both age groups.

Visual acuity was important in all studies because participants had to read target words from the computer screen (in both tested modalities). Obviously, visual acuity was particularly important in the two reading studies. Participants were asked to bring their glasses to the laboratory, if they had any, particularly the glasses they would use for looking at a computer screen. No further test was administered to measure participants' visual acuity. Large and easily legible fonts were used to minimize the chance that any subtle visual acuity differences affected the results. All participants also performed the DSS test (subpart of the Wechsler Adult Intelligence Test, 2004). Scores on the DSS test have been shown to exhibit strong correlations with measures involving perceptual speed or processing speed (Salthouse, 2000; Hoyer *et al.*, 2004). Mean substitution time needed per symbol was 1.83 s/symbol (SE=0.06) for the elderly and 1.35 s/symbol (SE=0.03) for the young participants. This should be corrected for motor speed (writing time), which was 0.93 s/symbol (SE=0.06) for the elderly and 0.71 s/symbol (SE =0.03) for the young participants. The corrected coding time (substitution time minus writing time) was then 0.90 s/symbol (SE=0.05) for the elderly and 0.64 s/symbol (SE =0.04) for the young participants.

Vocabulary knowledge of all participants was tested by way of the vocabulary subpart of the Groningen Intelligentie Test (Luteijn and van der Ploeg, 1983). Out of 20 multiple choice synonym questions, mean score correct was 18 (SE =0.2) for the elderly and 15 for the young (SE=0.4) participants. Better vocabulary scores for the elderly are often found in studies in which these two age groups are compared.

C. Material

The four conditions (two rates × two compression types) were distributed over 120 test sentences according to a Latin

square design. This design required four different stimulus lists to which the participants were randomly assigned.

The Dutch language material consisted of 120 disyllabic nouns as targets for detection: 60 with lexical stress on the initial and 60 with stress on the final syllable. The targets were embedded in sentences that varied from 7 to 15 words in length. The sentences were news bulletin sentences that were changed such that the noun would fit the sentence. The target words were possible but not fully predictable continuations of the sentence, e.g., *He did not pay enough attention to the traffic around him*, and *But there too the supply was increasingly more extensive and expensive* (target words underlined). The target's position in the sentence ranged from the second word in the sentence to the last word. Position in the sentence was not manipulated systematically. Additionally, the sentence materials comprised 60 overlap filler sentences in which the target had word-initial phoneme overlap with one of the words in the sentence (e.g., target being *cirkel* "circle" and sentence contains the word *circus*). There were also 40 miss filler sentences in which the target simply did not occur. One young male native speaker of Dutch read out all the materials at a normal speech rate (mean rate was 5.7 syllable/s). He was asked to place a sentence accent on the target word in the target sentences. The recording was made with an Audio-Technica AT 8410 microphone on a digital audio tape. All sentences were stored as separate files and downsampled to 32 kHz. Overall loudness of the audio files was equalized to 72 dB. Time compression was applied to the original recordings as discussed in Sec. II A.

D. Procedure

Testing was done in the Utrecht Institute of Linguistics Phonetics Laboratory in silent booths. Participants first performed the auditory word detection experiment. Then the diagnostic part was administered: the audiogram, the DSS test, and the vocabulary test. Then there was a short break after which they participated in the reading (word detection) experiment. Lastly, they performed the self-paced reading test.

In the auditory word detection study, the target words were presented visually in a large black font (Helvetica 80) on the computer screen 1200 ms before the auditory sentence started. The black target word was centered on the computer screen against a light gray background and remained on the screen during sentence presentation. Participants were instructed to respond as fast and as accurately as possible once they detected the pre-assigned visual target in the auditory sentence by pressing a button on the button box (with their dominant hand). Following sentence offset, participants still had 2 s to respond if they had not already done so, before the new visual target for the next sentence was presented. None of the participants reported feeling rushed by the pace of the experiment.

The auditory materials were presented binaurally over closed headphones (Beyer Dynamic DT 250) at a mean level of 81 dB SPL (for all listeners). Listeners were first presented with a practice part after which they could still ask questions if anything was unclear. They were then presented with the material, blocked by speech rate (first the sentences at 1.5 times the original rate, then the even faster sentences). Within each speech rate block, items made with either type of compression were presented in random order. Order of presentation of test and filler sentences was randomized for each participant.

As a check whether participants had actually paid attention to the content of the sentences, a sentence recognition test was administered to them immediately after the listening study. Participants were presented with a list of 16 sentences: 8 of which had been presented in the listening study and 8 similar sentences which had not been presented. They were asked to tick the sentences they thought they had just heard. Elderly participants on average correctly remembered 6.1 (SE=0.2) out of 8 sentences; the young participants had an average score of 5.75 (SE=0.2). The mean number of falsely recognized sentences was 0.58 (SE=0.1) out of 8 for the elderly and 0.20 (SE=0.1) for the young participants.

TABLE I. Correct word detection rates in the different conditions for both age groups.

	Compressed to 1.5×normal rate (67%)		Compressed to 2×normal rate (50%)	
	PSOLA	Sampling	PSOLA	Sampling
Young	99	100	99	98
Elderly	100	100	98	92

E. Results

The overall aim of the present study was to investigate, within a group of elderly listeners, which auditory and nonauditory measures determine one's ability to keep up with a fast speech rate. Before the results of the listening study will be related to background measures and performance in the reading experiments, the results of the listening study are presented and discussed as such for the two age groups to show that the increased speech rate indeed affected the elderly listeners more than the young listeners. Thus, an age group×rate interaction was anticipated. Furthermore, an age group×compression type interaction was anticipated: the type of time compression involving acoustic artifacts was expected to be more detrimental for the elderly listeners given their generally poorer hearing level.

A word detection response was counted as correct when the participant had pressed the button after target onset (and before the time out). Responses that were given within 200 ms from target onset were also excluded (for the young listeners this last criterion excluded 8 out of 2379 observations and 2 out of 4667 observations for the elderly listeners). In Table I, correct detection rates of both participant groups are given for the different test conditions.

Table I shows that correct detection rates were always higher than 92%. False response rates to the overlap fillers were 5% and 4% in the elderly listener group at the fast (1.5 times normal rate) and fastest rate of speech (2 times normal rate), respectively. False response rate to the filler sentences which did not contain a word related to the target (the miss

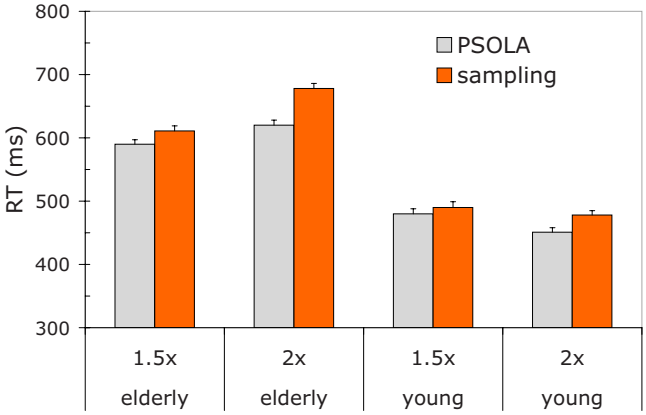


FIG. 2. (Color online) Mean response time (from target word onset) in the different rate conditions for both age groups. Error bars reflect standard errors.

condition) was 1% at both compression rates. False response rates for the young listeners were somewhat higher: 10% at both compression rates for the overlap fillers and 5% at both rates in the miss condition. These hit and false alarm rates indicate that the elderly listeners did not press the button randomly, nor did the younger listeners: collapsed over rates and conditions, d' (correct hits, relative to false alarms in overlap and miss conditions) for the elderly listeners was 3.73 and 4.50 for the young listeners.

Detection time was measured from target word onset. Mean average detection time in the different conditions is presented in Fig. 2.

The correct *RT* observations ($N=7036$) were logtransformed. Rather than by means of two analyses of variance, one having participants as a random effect and one with items as a random effect, mixed-effect modeling was used to analyze the data. The mixed-effect method allows one to acknowledge that multiple crossed random factors (participants and items) affect performance in psycholinguistic designs as the one used here (Pinheiro and Bates, 2000). Similar mixed-effect approaches were taken in acoustic measurement studies of speech corpus data in which speaker and item are multiple random factors (Pluymaekers *et al.*, 2005; Ernestus *et al.*, 2006; Quené, 2008). Mixed-effect modeling is argued to be superior to repeated measures or univariate analysis of variance (ANOVA) in a number of tutorial articles (Baayen *et al.*, 2002; Quené and van den Bergh, 2004, 2008; Baayen *et al.*, 2008).

The log *RT* data were analyzed here to test for the effects of age group, rate, compression type (PSOLA/sampling), and stress position and possible interactions between them. For all significant effects and interactions, estimated coefficients (β's), with standard errors for β in brackets, and p-values are reported.[1] The results showed a main effect of age group [estimated coefficient $=0.092(0.029)$, $p<0.001$]: the elderly participants were generally slower than the young. There was a main effect of rate [$\beta=-0.025(0.009)$, $p<0.005$], and more importantly, there was a significant age×rate interaction [$\beta=0.045(0.010)$, $p<0.001$]. As can be seen in Fig. 2, whereas the young listeners responded faster in the 2 times normal-rate condition than in the 1.5 times normal-rate condition, the reverse is true for the elderly. It is important to note that if one measures *RT*'s from target word *offset*, rather than from target *onset*, all participants have increased *RT*'s when speech rate is increased further (measured from target word offset, and collapsed over compression types, the young participants have a mean *RT* of 199 ms in the 1.5 times normal rate and 249 ms in the 2 times normal-rate condition, whereas the elderly participants go from 314 to 432 ms in the fastest condition). There was no overall effect of compression type [$\beta=0.003(0.009)$, *n.s.*]. However, there was a significant interaction between rate and compression type [$\beta=0.030(0.012)$, $p<0.05$]: the *RT* advantage of PSOLA over the sampling method was larger at the fastest (2 times original) rate than at the intermediate rate (1.5 times original). However, there was no interaction between age and compression type, nor a three-way interaction between age, compression type, and rate. Lastly, the effect of stress position of the target items was not significant overall, nor its interaction with age group [$\beta=-0.014(0.010)$, *n.s.*].

A very similar pattern of significant main effects and interactions was found when *RT* results were analyzed from target offset, rather than from target onset.

Accuracy of the responses in the different conditions was analyzed as well (note that accuracy was high overall). This was done with mixed-effect modeling for binomial data (responses being either correct or false). Quené and van den Bergh (2008) showed that mixed-effect logistic models provide a more conservative and appropriate method to analyze binomial data than the traditional practice of computing proportions correct over subjects or items. The effects of age group, rate, compression type, and stress position were tested and their interactions (estimated coefficients are reported, with their standard errors in brackets, and p-values). The mixed-effect model showed a significant effect of age group [the elderly having lower hit rates than the young: $\beta=-1.439(0.382)$, $p<0.001$]. The main effect of rate [$\beta=-1.954(0.503)$, $p<0.001$] was significant: detection accuracy was decreased at the fastest speech rate. There was no main effect of compression type [$\beta=0.160(0.673)$, *n.s.*], but the interaction between rate and compression type was significant [$\beta=-1.771(0.714)$, $p<0.05$]. Performance of both age groups was more affected by the acoustic artifact at the faster than at the intermediate (1.5 times normal) rate of speech. None of the other effects or interactions (including the interaction between age group and rate) was significant.

Thus, in the detection rate analysis and in the response time analysis, the effect of compression type was more disrupting at the faster rates, and this was found for both age groups.

The two reading studies will now be presented first, before the fast speech results are related to reading performance and to the background measures.

III. VISUAL WORD DETECTION STUDY

A. Participants

The participants in the two reading studies (visual word detection study and self-paced reading) were the same as

those who participated in the listening study. Due to technical failure during one test session, reading performance of 1 of the 40 elderly participants who had completed the listening experiment was not available. The results reported below are based on performance of 40 elderly listeners (thus, one replacement participant was included from the same elderly group).

B. Materials

One major complication in comparing rate in the auditory and visual modalities is that auditory presentation rate can always be related to "normal" or original speech rate: the rate at which the speaker uttered the sentence. In the visual modality, normal reading rate is up to each individual reader. On the basis of several studies on reading performance of elderly adults with sequential (word-by-word) presentation methods (Stine, 1990; Speranza *et al.*, 2000; Hartley *et al.*, 1994), normal presentation duration was estimated to be 400 ms per word. This estimation was on the safe side since this rate was to be speeded up.

The visual word detection study had an analogous set-up and similar sentence materials as the auditory word detection study. Sixty bisyllabic target words, always morphologically simplex nouns, were embedded in sentences in which they were not completely predictable from the sentence context. As in the listening experiment, half of the target items had initial stress; the other half had final stress. The sentence material was presented word-by-word: each individual word was displayed on the screen equally long. Given that normal reading rate was estimated to be 400 ms per word, the two rate conditions were made by speeding up presentation duration to 267 (67% rate condition) or 200 ms (50% rate condition) per word. Sentences were 8–19 words long. None of the words in the sentences had more than 12 letters. The two presentation rate conditions were distributed over the 60 test sentences according to a Latin square design. This design required two different stimulus lists to which participants were randomly assigned.

As in the auditory experiment, there were also filler trials: overlap fillers containing a noun that showed word-initial overlap with the target noun and miss fillers not containing a word that (partly) matched the target. The ratio of overlap and miss fillers to test sentences was identical to that in the auditory experiment: in addition to the 60 test sentences, there were 40 overlap trials and 26 miss trials.

C. Procedure

As in the auditory word detection experiment, participants were first presented with a visual target word, 1000 ms before the first word of the sentence was shown. The target word had a different font (Courier bold which has serifs; 50 point) and color (dark green) than the words of the sentence, and was in upper case, unlike the words of the sentence. This was done to make the target word as distinctly different as possible from the sentence such that participants would really have to read the sentence and not be able to detect the target word purely on the basis of superficial "image" matching. The sentence words were in a black Helvetica 50 point

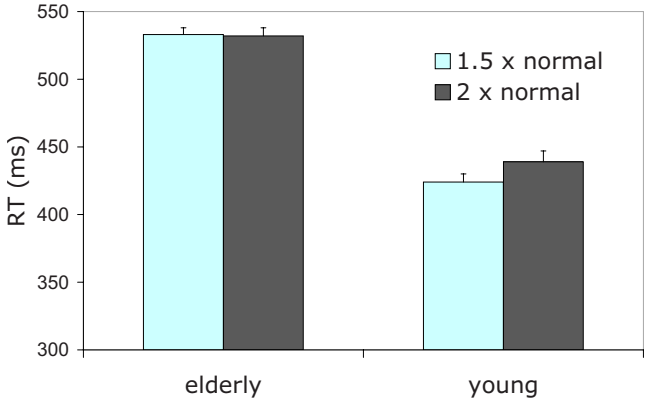


FIG. 3. (Color online) Mean response time in visual word detection experiment at both rates of presentation for both age groups (error bars reflect standard errors).

letter (this font has no serifs) on the screen: the words appeared in a white box against a light gray screen background. The target word remained on the screen during sentence presentation, clearly distinct from and under the box in which the words of the sentence were presented. Participants had 2 s following sentence offset to respond before the new visual target for the next sentence was presented. They were instructed to respond as fast and as accurately as possible once they detected the pre-assigned target in the sentence by pressing a button on the button box (with their dominant hand). They were informed that there would be sentences not containing the target word or with words similar to the target word. Participants were first presented with a practice part of eight sentences (containing both four test and four filler sentences in which the target would not appear: two misses and two with an overlapping word) after which they could still ask questions if anything was unclear. They were then presented with the material, blocked by speech rate (first the moderately fast sentences, then the faster sentences). Order of presentation of test and filler sentences was randomized for each participant.

D. Results

As in the auditory study, the increased rate of visual presentation was expected to have a more pronounced effect on the elderly than on the young participants. However, the two rates of visual presentation in the word detection study were hardly reported as challenging by the elderly participants. As in the listening study, response time was computed from target onset. Target detection times are given in Fig. 3 for both age groups and both presentation rates (1.5 and 2 times the normal rate, or compression to 67% and 50% of the normal duration, respectively). Correct detection rates were 99% for the elderly (at both presentation rates) and 100% for the young participants (at both rates).

The (correct) detection times ($N=3571$) were logtransformed and analyzed with linear mixed-effect models to test the effects of age group, presentation rate, and stress position of the target item. As in Sec. II E, estimated coefficients are reported, plus standard error for the estimate between brackets, and p-values. Elderly participants generally responded

more slowly than the young listeners [age group effect: $\beta = 0.099(0.024)$, $p < 0.001$], but there was no overall effect of rate of presentation [$\beta = 0.005(0.007)$, *n.s.*]. Responses to initially stressed items (word-initial stress being the default stress pattern in Dutch) were overall faster than to finally stressed items [main effect of stress position $\beta = 0.029(0.010)$, $p < 0.005$], and this effect was slightly more pronounced for the young than the elderly listeners [$\beta = -0.023(0.009)$, $p < 0.05$].

There was no specific hypothesis concerning the effect of stress position for the different age groups. Nevertheless, these reading results show that the elderly participants are less affected by stress position of the lexical item than the young participants. Increased exposure over the years to words that do not match the default stress pattern might explain this result.

The age group × presentation rate interaction was not significant [$\beta = 0.002(0.009)$, *n.s.*]: if anything, the elderly participants as a group seemed to show less influence of the increased presentation rate than the young (cf. Fig. 3). None of the other interactions reached significance.

IV. SELF-PACED READING

A. Participants

Participants were the same as those who participated in the auditory word detection study and in the visual word detection study (cf. Sec. III A).

B. Stimuli and procedure

In the self-paced reading experiment, participants were asked to read a short folk story (about a monkey) at their own preferred pace. With each press on a response button of a button box, the next word of the sentence would appear (or the first word of the next sentence). Even though each word disappeared before the next word was presented, the sentence was presented "from left to right:" each new word appeared in the position in which it would have been if the sentence had been presented as a whole at once. The dependent variable is reading time per word, or the pause between consecutive button presses. The words were presented in a white Helvetica font (size 30) against a black background (rectangle in which the sentence would fit). This sentence box was positioned just above the middle of the screen.

Participants received an oral instruction how to perform the task. By way of a practice set, the first sentence pair introduced the task to the participants once more: "You are going to read a short story now." "You decide what the tempo will be by pressing the button on the button box." If the participants had no more questions, they proceeded to the test part. The short story consisted of ten sentence pairs (the introductory practice sentence pair excluded), each of which consisted of two sentences. These sentences varied in length between 7 and 18 words.

C. Results

Overall slower (i.e., longer) reading times were expected for the elderly participants than for the young participants. In

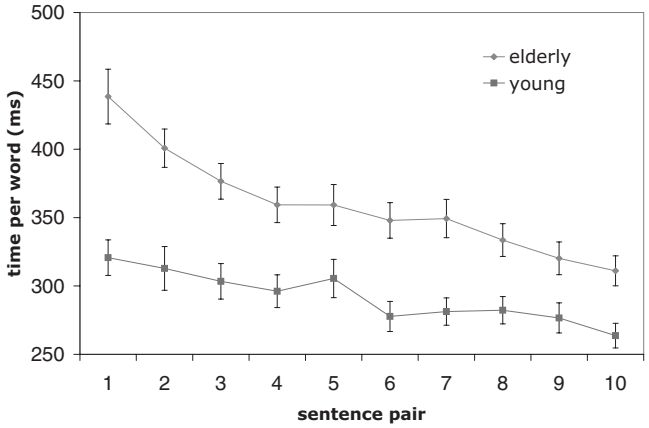


FIG. 4. Mean reading time per word per sentence pair for both age groups (error bars represent 1 SE).

Fig. 4 mean reading time per word is displayed for the two age groups as a function of sentence pair number. Because the story makes more and more sense toward the end, mean reading time per word shows a general downward trend. An ANOVA (by subjects) was carried out on these data because mixed-effect models (with both subjects and items as random factors) did not seem appropriate here. The (logtransformed) reading time per word data showed a main effect of age group [$F(1,58) = 4.85$, $p = 0.032$] and a main effect of sentence pair number [$F(9,49) = 43.07$, $p < 0.001$]. As Fig. 4 suggests, the interaction between age group and sentence pair is also significant [$F(9,49) = 6.21$, $p < 0.001$]: the difference in reading time between young and elderly listeners becomes smaller over the course of the short story.

The purpose of this self-paced reading experiment was to investigate a relationship between self-preferred reading rate and performance in the listening study. Therefore, each individual participant's reading pace to be entered as a predictor for fast speech performance was not computed over the entire story, but was restricted to the first two sentences because these set up a new semantic context frame. The later sentences introduce new content as well, but also refer back to concepts and persons introduced earlier. This fact that the first two sentences of the story introduce more new persons and concepts than later sentences makes them more similar to the sentence materials in the two word detection studies where every sentence is completely unrelated to the previous one and sets up its own new context. Mean reading time per word (computed over the first two sentence pairs) was 317 ms per word for the young adults (SE=10) and 419 ms for the elderly adults (SE=10, but note that 4 out of 160 observations were discarded because they were more than three SEs removed from the mean). An ANOVA on (logtransformed) reading times showed that this reading time difference between the age groups was significant [$F_1(1,58) = 13.67$, $p < 0.001$; $F_2(1,3) = 155.00$, $p = 0.001$]. Computed over the entire story, mean reading time per word was 292 ms for the young adults (SE=4) and 359 ms for the elderly adults (SE=4).

V. COMBINED RESULTS

The following three reading measures were considered potential predictors for fast speech performance: partici-

TABLE II. Correlation matrix with correlation coefficients (Pearson's) between all predictor pairs for the group of 39 elderly participants. Asterisk notation " * ": $p<0.05$.

	Hearing acuity	Age	Digit symbol	Vocab	Self-paced reading	log *RT* (visual word detection)	Rate effect (visual word detection)
Hearing acuity		0.27	−0.17	−0.05	−0.10	0.31	0.35*
Age	0.27		−0.08	0.25	0.14	0.11	0.12
Digit symbol	−0.17	−0.08		−0.19	0.10	−0.08	0.04
Vocabulary score	−0.05	0.25	−0.19		0.25	0.09	−0.16
Self-paced reading	−0.10	0.14	0.10	0.25		0.05	0.15
Overall log *RT* in visual word detection study	0.31	0.11	−0.08	0.09	0.05		0.34*
Rate effect in visual word detection	0.35*	0.12	0.04	−0.16	0.15	0.34*	

pants' self-paced reading time (log time per word), overall target detection time (log) in the visual word detection study, and the effect of increased reading rate. Even though there was no overall effect of increased presentation rate in the visual word detection study, there were individual differences in the effect of presentation rate: the effect was computed for each of the elderly participants by dividing their mean (log) response time in the fastest presentation rate by their mean log *RT* in the slower presentation rate. As said, due to technical failure, this combination of reading and listening data applied to the data of 39 out of 40 elderly individuals. The other background measures were age, vocabulary score, hearing acuity, and corrected DSS score. For each listener, mean hearing loss in the better ear was computed by averaging pure-tone averages over 1, 2, and 4 kHz. Vocabulary performance was entered as the proportion of correct responses. Age was entered after subtraction of the minimum age of 65 years (because values of all other factors and predictors in the model are estimated for the minimum age value which would be far outside the actual age range if minimum age is not subtracted; cf. Snijders and Bosker, 1999 for the issue of centralization of predictor values). With respect to DSS test performance, each individual's corrected coding time (time required to recode one symbol) was entered.

Correlations among the different background measures were computed within the group of 39 elderly to address the issue of collinearity. If they are highly correlated, it does not make sense to consider them jointly in a multiple regression model of elderly participants' fast speech performance. Table II presents a correlation matrix with Pearson correlation coefficients between all predictor pairs (one asterisk denoting that $p<0.05$).

Given the weak (if any) correlations among these background measures, they were all entered into the analyses. The listening results were analyzed by using linear mixed-effect models with subject and item as crossed random effects (Pinheiro and Bates, 2000; Quené and van den Bergh, 2004; Bates and Sarkar, 2005). An additional advantage of mixed-effect models, over the advantage that multiple random effects can be considered in one analysis, is that both categorical and discrete predictors can be included in a single model (e.g., a manipulated factor, such as compression type, and individual characteristics, such as hearing acuity). Several models were fitted to the binomial detection accuracy data (using logistic regression, valid correct and incorrect responses made up 4678 observations), and several models were fitted to the detection time data (correct responses only, $N=4549$). It is important to note that in linear mixed-effect models, the order in which the predictors are entered does not influence their predictive value. The models evaluate whether coefficients are significantly different from zero in a model that contains all other predictors.

A. Detection rate

In Table III, an overview is given of four tested models. Model 0 is the simplest linear mixed-effect model which only contains as fixed factors the manipulated factors compression rate, compression type, and their interaction, and takes participant and item as random effects (as will be the case in all later models as well). In model 1, hearing acuity is added to the fixed part see how its inclusion improves the previous model. In model 2, self-paced reading time is added to model 1, and lastly, digit symbol coding time is added to the previous model in model 3. Table III also contains an evaluation of the four consecutive models in terms of goodness of fit. This is done by the likelihood ratio test, rather than by a computation of R^2 for each of the models because in mixed-effect models different sources of variation are modeled jointly (fixed and random effects). If R^2 were to be computed, it would not only cover the variance explained by the fixed effects (as is normally the case in multiple regression analysis, e.g., Jerger *et al.*, 1991) but also that of the random effects. Table III also shows how inclusion of the background measure predictors in the later, more complex models leaves less and less random between-subject variance (relative to model 0).

The log likelihood ratio test takes log likelihood, a measure of the model's goodness of fit, for the simpler model (containing less predictors) and compares it with the log likelihood for the more complex model with more parameters. The difference between the two log likelihoods, multiplied by 2, follows a chi-square distribution with the difference in number of parameter as number of degrees of freedom. In other words, the output of this likelihood ratio test determines whether inclusion of more parameters is justified. In Table III, each later model is evaluated relative to the previous model, and each model has a better fit than the previous one. Asterisks denote which of the coefficients are significant within that model.

This model comparison shows that model 3 contains

TABLE III. Comparison of several mixed-effect models of the accuracy data: estimated parameters (β coefficients) for the fixed part, variance of the two random effects, and evaluation characteristics (significant codes: " *** ": $p<0.001$, " ** ": $p<0.01$, " * ": $p<0.05$, and " . ": $p<0.1$).

		Model 0	Model 1	Model 2	Model 3
Fixed effects	Compression rate	−2.26***	−2.23***	−2.25***	−2.27***
	Compression type	−0.22	−0.22	−0.24	−0.25
	Compression rate × compression type	−1.48.	−1.50.	−1.48.	−1.48.
	PTA		−0.06***	−0.06***	−0.07***
	SPR (log reading time)			−3.23*	−4.27*
	Digit symbol coding time				−1.05*
Random	Participant variance	1.06	0.50	0.38	0.27
	Item variance	5.43	5.46	5.54	5.59
Evaluation	Log likelihood	−413.08	−404.46	−401.55	−399.37
	Chi-square (compared to previous model)		17.242	5.84	4.35
	p-deviance (compared to previous model)		<0.001	<0.05	<0.05

predictor values that show significant effects over and above the manipulated factors. The other background measures (age, vocabulary score, mean log *RT* in reading, and the rate effect in reading) turned out not to predict fast speech performance.[2] Interaction terms between design factors and predictor variables were also added, but this did not improve the model.

Thus, the model with the best fit, model 3, showed the following effects (as in previous discussions of linear mixed-effect results for binomial data, estimated coefficients are reported plus their standard errors, and significance levels). Compression rate had a significant main effect on detection rate [estimated coefficient (henceforth β) is negative, meaning that the higher the speech rate, the lower the detection rate, compared to 1: $\beta=-2.270(0.604)$, $p<0.001$], but the main effect of compression type was not significant [$\beta=-0.250(0.750)$, *n.s.*]. The interaction between compression rate and compression type approached significance [$\beta=-1.477(0.797)$, $p=0.064$]. Hearing acuity [pure tone average (PTA)] significantly influenced detection rate: the more hearing loss, the lower the detection rate [$\beta=-0.067(0.012)$, $p<0.001$]. In addition, mean reading time per word (logtransformed, self-paced reading test) significantly influenced detection rate: the longer the reading time per word, the lower the detection rate [$\beta=-4.270(1.339)$, $p<0.01$].[3] Lastly, performance in the DSS test (as a measure of information processing rate) also significantly influenced performance: the longer one took to recode a digit to a symbol, the lower the detection rate [$\beta=-1.050(0.471)$, $p<0.05$]. Thus, hearing acuity, self-preferred reading rate, and information processing rate (in this order of relative importance) predicted success of target detection in fast speech.

B. Response time

Several linear mixed-effect models were fitted to the log *RT* data of the fast speech study (4547 observations). The following measures did not predict response time (nor interacted with any of the factors): vocabulary score, age, and hearing acuity. Table IV lists the characteristics of three

TABLE IV. Comparison of several mixed-effect models of the log *RT* data: estimated parameters (β coefficients) for the fixed part, variance of the two random effects, and evaluation characteristics.

		Model 0	Model 1	Model 2
Fixed effects	Compression rate	0.02***	0.02***	−0.81***
	Compression type	0.01*	0.01*	0.01*
	Compression rate × compression type	0.03***	0.03***	0.03***
	Digit symbol coding time		0.002	−0.01
	log *RT* in visual word detection study		0.89***	0.89***
	Rate effect in visual word detection		−1.09	−1.48
	Self-paced reading time		0.22.	0.22.
	Compression rate × digit symbol coding time			0.03**
	Compression rate × rate effect in visual detection			0.81***
Random	Participant variance	0.01	0.007	0.007
	Item variance	0.003	0.003	0.003
Evaluation	Log likelihood	3478	3491	3501
	Chi-square (compared to previous model)		24.42	20.73
	p-deviance (compared to previous model)		<0.001	<0.001

models: again moving from the simplest model (model 0) that only contains the interaction between the design factors, to a model that contains more predictors (model 1), to a model that contains interactions between factors and predictors (model 2).

The evaluation part of Table IV shows that adding background measures (model 1) improved the model, relative to the simplest model (model 0). The asterisk notation shows that model 1 is better than model 0 only because two, rather than four, background measures (namely, log *RT* in reading and the self-paced measure) were significant: the other two background measures were not significant predictors in model 1. The latter two background measures turn out to be important in their interaction with the compression rate effect in model 2, however. One should also note that the value of the compression rate coefficient was positive in models 0 and 1, meaning that generally, response times became longer in the faster rate condition (as was shown in Fig. 2). The direction of the rate effect is reversed in model 2: now, the faster the speech rate, the shorter the response times ($\beta=-0.809$), as was also seen for the young adults. Thus, the best-fitting model (model 2) showed a significant main effect compression rate on log *RT* [$\beta=-0.809(0.227)$, $p<0.001$], and a main effect of compression type [$\beta=0.011(0.004)$, $p<0.05$], indicating that response times were longer in the sampling conditions than in the PSOLA time-compressed conditions. This effect was stronger at the faster speech rate, as shown by a significant interaction between rate of speech and compression type [$\beta=0.032(0.006)$, $p<0.001$].

Mean detection *RT* in the visual word detection experiment was also a significant predictor: the longer one's detection time in reading, the longer one's detection time in listening [$\beta=0.892(0.179)$, $p<0.001$]. This reflects the shared components in the two analogous studies: first in terms of basic motor speed (how fast can this individual press the response button) and second in terms of speed of language processing and decision making. Self-paced reading time was another predictor that was very close to significance: the longer the reading time, the longer the detection time in listening [$\beta=0.218(0.119)$, $p=0.067$].[4] There was no overall effect of digit symbol coding time on *RT*'s, but it did interact with the compression rate effect [$\beta=0.027(0.010)$, $p<0.01$]. Whereas listeners generally have shorter response times in the fastest rate condition (relative to the 1.5 times normal-rate condition, cf. the negative coefficient for compression rate), this effect is the more counteracted, the longer time one needed to recode a digit to a symbol. A similar interaction was seen between compression rate and the rate effect in the visual word detection experiment, which in itself did not have a main effect on detection *RT*. Again, whereas *RT*'s generally get shorter in the fastest speech condition, this effect is counteracted more for those participants with a larger rate effect in reading [$\beta=0.810(0.227)$, $p<0.001$]. This indicates that participants who responded relatively slowly at the faster rate of visual presentation did not speed up in the listening study either when presented with the faster rate of speech.

Thus, the model fitted to the response times mainly showed interesting predictors for rate effect size: the extent to which participants were affected (i.e., sped up or slowed down) by increased rates of speech. Importantly, hearing acuity was not related to increased fast speech difficulty as measured by response time, but keep in mind that hearing level was the most important measure in explaining detection success. The model did show that increased difficulty with faster rates of speech was predicted by (a) the effect of presentation rate in the visual word detection experiment and (b) by information processing rate (DSS task performance).

VI. DISCUSSION AND CONCLUSION

The present study was set up to investigate why elderly listeners are more affected than young listeners when they are presented with fast, or more particularly, time-compressed speech. This is an old question that has given rise to a large number of studies (Gordon-Salant and Fitzgibbons, 1993, 1995, 1999, 2001; Letowski and Poch, 1996, Tun, 1998, Wingfield *et al.*, 1999; Schneider *et al.*, 2005). Even though many studies have acknowledged the contributions of hearing and cognitive factors to this age × rate interaction, it has been relatively difficult to quantify their relative contributions. The idea behind the present study was that if age-related general cognitive slowing and decreased information processing rate may be (partly) responsible for elderly listeners' increased difficulty with time-compressed speech, it makes sense to focus on this effortful process itself, rather than on just the result of it. Allowing elderly participants unlimited time to come up with a response may blur our conception of why perceptual processing itself has become more effortful. Moreover, the use of an on-line measure yields complementary accuracy and response time data.

The present study showed significant correlations between, on the one hand, hearing acuity and fast speech performance and, on the other hand, between individual presentation rate effects in word-by-word reading and in speech. Hearing loss clearly plays a primary role in explaining elderly listeners' problems with fast speech, as was evident from the detection rate analysis. Having to decode an impoverished speech signal requires more effort and thus takes more time, which becomes all the more problematic when speech rate is fast.

Note that the present data replicated the finding of Humes *et al.* (2007) that *overall* performance of the elderly participants was not affected by rate of visual presentation. Clearly, increased presentation rate affects cognitive processing more in the auditory modality than in the visual modality. This ties in with the suggestion by Humes *et al.* (2007) that high-frequency hearing loss is associated with central auditory problems, such as poorer spectral and temporal resolution, over and above the loss of audibility. In their results, high-frequency hearing loss turned out to be a significant predictor of fast speech performance, even after spectral shaping had been applied to compensate for loss of audibility. Nevertheless, despite the fact that overall performance of the elderly participants was not affected by rate of presentation, the present focus on processing time has resulted in two important findings. Even though correlations in performance across modalities have been found before (Zekveld *et al.*,

2007; Humes *et al.*, 2007), the present study is the first to show correlations between *rate effects* across modalities. Furthermore, the present results provide support for a direct relationship between individual differences in processing speed (DSS) and fast speech perception (cf. Sommers, 2005). It is not surprising that performance in the two reading tasks, being language measures, turned out to be better predictors than the more general cognitive measure of information processing rate (DSS).

The combination of correlations between fast speech performance and self-paced reading rate (detection accuracy analysis) and between individual rate effects in the two analogous word detection studies (*RT* analysis) provides support for an underlying problem with fast language processing: language being a particularly complex type of information to process. At first sight, it seems that the present study's "fast rates" of visual presentation were not fast or challenging enough. Note, however, that normal reading rate for the elderly listeners was estimated at 400 ms per word, which turned out to be remarkably close to elderly participants' mean self-preferred reading time (419 ms per word) in the first sentences of the folk story.

Self-paced reading may have given a more reliable insight into individual speed of language processing than the visual target detection study. It seems unlikely that participants ignored the content of the sentence while detecting pre-assigned target words: the target item that elicited the fastest detection time may have been more predictable from its preceding sentence context than average *(All of a sudden tapped he with a pencil on the window).* However, there was no check on whether all participants actually remembered some of the visually presented sentences afterward, which leaves the possibility that (some) participants only scanned the sentences for an orthographic match to the target. Even though self-paced reading had the better predictive power of the two for fast speech performance (cf. the individual detection rate analysis), the related effects of speeding in the visual and auditory modalities provide support that there is a "general cognitive" component to the age × rate interaction in listening to fast speech.

This brings us back to the claim by Schneider *et al.* (2005) that elderly listeners' problems with speeded speech should be attributed mainly to their inability to cope with the acoustic artifacts involved in methods of time compression, rather than to cognitive slowing. The present results did not show a clear age group × compression type interaction, but Schneider *et al.* (2005) may be right that, in general, any acoustic artifact is more disrupting for elderly than young listeners. Yet, the rate correlations across modalities provide solid evidence that there is more to the age × rate interaction than acoustic artifacts. Future research could investigate the potential benefits of non-linear over linear time compression, given their finding that elderly listeners were not more affected if speeding was accomplished by only removing silences and steady-state portions of the signal.

In conclusion, the present study has provided direct evidence that auditory and cognitive factors underlie elderly listeners' problems keeping up with fast rates of speech. Hearing loss, combined with central auditory problems that may be linked to high-frequency hearing loss, is certainly the primary cause of the problem. Nevertheless, be it at a more subtle level, general information processing problems also affect rate of language processing in auditory and non-auditory domains.

ACKNOWLEDGMENTS

The Dutch Organisation for Scientific Research (NWO) is gratefully acknowledged for funding this research (Grant No. 275-75-004). I would like to thank Sieb Nooteboom for his comments on earlier drafts of this study and Mirjam Ernestus and Hugo Quené for their statistical advice.

[1] *p*-values (obtained by generating 10 000 simulations of the data sample) were computed with an ancillary function pvals.fnc defined in the LANGUAGER package. This function can be applied to the fitted model (http://cran.r-project.org, see Baayen, 2008).

[2] When raw age was in the model (instead of age minus 65) it was not a significant predictor either.

[3] A somewhat weaker, but still significant, predictor of detection rate is individual self-paced reading time per word averaged over the entire story, rather than averaged over only the first two sentences of the story [β = −2.899(1.187), p=0.015].

[4] If self-paced reading time is averaged over the entire story, rather than over the first two sentences, it is not a significant predictor [β =0.119,(0.098), *n.s.*].

Baayen, R. H. (**2008**). *Analyzing Linguistic Data: A Practical Introduction to Statistics* (Cambridge University Press, Cambridge).

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**, 390–412.

Baayen, R. H., Tweedie, F. J., and Schreuder, R. (**2002**). "The subjects as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon," Brain Lang. **81**, 55–65.

Balota, D. A., and Duchek, J. M. (**1988**). "Age-related differences in lexical access, spreading activation, and simple pronunciation," Psychol. Aging **3**, 84–93.

Bates, D., and Sarkar, D. (**2005**). "The lme4 library," available online at http://lib.stat.cmu.edu/R/CRAN/ (Last viewed April, 2008).

Covell, M., Withgott, M., and Slaney, M. (**1998**). "Mach1: Nonuniform time-scale modification of speech," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Seattle, pp. 493–496.

Dubno, J. R., Dirks, D. D., and Schaefer, A. B. (**1987**). "Effects of hearing loss on utilization of short-duration spectral cues in stop consonant recognition," J. Acoust. Soc. Am. **81**, 1940–1947.

Ernestus, M., Lahey, M., Verhees, F., and Baayen, R. H. (**2006**). "Lexical frequency and voice assimilation," J. Acoust. Soc. Am. **120**, 1040–1051.

Federmeier, K. D., McLennan, D. B., De Ochoa, E., and Kutas, M. (**2002**). "The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study," Psychophysiology **39**, 133–146.

Federmeier, K. D., Schwartz, T. J., Van Petten, C., and Kutas, M. (**2003**). "Sounds, words, sentences: Age-related changes across levels of language processing," Psychol. Aging **18**, 858–872.

Fink, M., Churan, J., and Wittmann, M. (**2006**). "Temporal processing and context dependency of phoneme discrimination in patients with aphasia," Brain Lang. **98**, 1–11.

George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., and Houtgast, T. (**2007**). "Auditory and nonauditory factors affecting speech reception in noise by older listeners," J. Acoust. Soc. Am. **121**, 2362–2375.

Glasberg, B. R., and Moore, B. C. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

Gordon-Salant, S., and Fitzgibbons, P. (**1993**). "Temporal factors and speech recognition performance in young and elderly listeners," J. Speech Hear. Res. **36**, 1276–1285.

Gordon-Salant, S., and Fitzgibbons, P. (**1995**). "Recognition of multiply

degraded speech by young and elderly listeners," J. Speech Hear. Res. **38**, 1150–1156.
Gordon-Salant, S., and Fitzgibbons, P. (**1999**). "Profile of auditory temporal processing in older listeners," J. Speech Lang. Hear. Res. **42**, 300–311.
Gordon-Salant, S., and Fitzgibbons, P. J. (**2001**). "Sources of age-related recognition difficulty for time-compressed speech," J. Speech Lang. Hear. Res. **44**, 709–719.
Hartley, J. T., Stojack, C. C., Mushaney, T. J., Annon, T. A. K., and Lee, D. W. (**1994**). "Reading speed and prose memory in older and younger adults," Psychol. Aging **9**, 216–223.
Hoyer, W. J., Stawski, R. S., Wasylyshyn, C., and Verhaeghen, P. (**2004**). "Adult age and digit symbol substitution performance: A meta-analysis," Psychol. Aging **19**, 211–214.
Humes, L. E. (**2005**). "Do 'auditory processing' tests measure auditory processing in the elderly?," Ear Hear. **26**, 109–119.
Humes, L. E., Burk, M. H., Coughlin, M. P., Busey, T. A., and Strauser, L. E. (**2007**). "Auditory speech recognition and visual text recognition in younger and older adults: Similarities and differences between modalities and the effects of presentation rate," J. Speech Lang. Hear. Res. **50**, 283–303.
Janse, E. (**2003**). *Production and Perception of Fast Speech* (Landelijke Onderzoeksschool Taalwetenschap, Utrecht).
Janse, E. (**2004**). "Word perception in fast speech: Artificially time-compressed vs. naturally produced fast speech," Speech Commun. **42**, 155–173.
Jerger, J., Jerger, S., and Pirozzolo, F. (**1991**). "Correlational analysis of speech audiometric scores, hearing loss, age, and cognitive abilities in the elderly," Ear Hear. **12**, 103–109.
Letowski, T., and Poch, N. (**1996**). "Comprehension of time-compressed speech: Effects of age and speech complexity," J. Am. Acad. Audiol. **7**, 447–457.
Luteijn, F., and van der Ploeg, F. A. E. (**1983**). *Handleiding Groninger Intelligentietest (Manual Groningen Intelligence Test)* (Swets and Zeitlinger, Lisse, The Netherlands).
McCoy, S. L., Tun, P. A., Cox, C. L., Colangelo, M., Stewart, R. A., and Wingfield, A. (**2005**). "Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech," Q. J. Exp. Psychol. A **58**, 22–33.
Moulines, E., and Charpentier, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun. **9**, 453–467.
Myerson, J., Ferraro, F. R., Hale, S., and Lima, S. D. (**1992**). "General slowing in semantic priming and word recognition," Psychol. Aging **7**, 257–270.
Nittrouer, S., and Boothroyd, A. (**1990**). "Context effects in phoneme and word recognition by young children and older adults," J. Acoust. Soc. Am. **87**, 2705–2715.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing: I. Intelligibility differences between clear and conversational speech," J. Speech Hear. Res. **28**, 96–103.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing: II. Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1989**). "Speaking clearly for the hard of hearing: III. An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech," J. Speech Hear. Res. **32**, 600–603.
Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise," J. Acoust. Soc. Am. **97**, 593–608.
Pinheiro, J. C., and Bates, D. M. (**2000**). *Mixed-Effects Models in S and S-Plus* (Springer, New York).
Pluymaekers, M., Ernestus, M., and Baayen, R. H. (**2005**). "Lexical frequency and acoustic reduction in spoken Dutch," J. Acoust. Soc. Am. **118**, 2561–2569.
Quené, H. (**2008**). "Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo," J. Acoust. Soc. Am. **123**, 1104–1113.
Quené, H., and van den Bergh, H. (**2004**). "On multi-level modeling of data from repeated measures designs: A tutorial," Speech Commun. **43**, 103–121.
Quené, H., and van den Bergh, H. (**2008**). "Examples of mixed-effects modeling with crossed random effects and with binomial data," J. Mem. Lang. **59**, 413–425.
Rosen, S. (**2003**). "Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything?" J. Phonetics **31**, 509–527.
Salthouse, T. A. (**2000**). "Aging and measures of processing speed," Biol. Psychol. **54**, 35–54.
Schneider, B. A., Daneman, M., and Murphy, D. R. (**2005**). "Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing?" Psychol. Aging **20**, 261–271.
Snijders, T. A. B., and Bosker, R. J. (**1999**). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Sage, London).
Sommers, M. S. (**2005**). in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Oxford), pp. 469–493.
Sommers, M. S., and Danielson, S. M. (**1999**). "Inhibitory processes and spoken word recognition in young and elderly adults: The interaction of lexical competition and semantic context," Psychol. Aging **14**, 458–472.
Sommers, M. S., and Humes, L. E. (**1993**). "Auditory filter shapes in normal-hearing, noise-masked normal, and elderly listeners," J. Acoust. Soc. Am. **93**, 2903–2914.
Spehar, B., Tye-Murray, N., and Sommers, M. (**2004**). "Time-compressed visual speech and age: A first report," Ear Hear. **25**, 565–572.
Speranza, F., Daneman, M., and Schneider, B. A. (**2000**). "How aging affects the reading of words in noisy backgrounds," Psychol. Aging **15**, 253–258.
Stine, E. A. L. (**1990**). "On-line processing of written text by younger and older adults," Psychol. Aging **5**, 68–78.
Tallal, P. (**1980**). "Auditory temporal perception, phonics and reading disabilities in children," Brain Lang. **9**, 182–198.
Tun, P. A. (**1998**). "Fast noisy speech: Age differences in processing rapid speech with background noise," Psychol. Aging **13**, 424–434.
Tun, P. A., Wingfield, A., Stine, E. A. L., and Mecsas, C. (**1992**). "Rapid speech processing and divided attention: Processing rate vs. processing resources as an explanation of age effects," Psychol. Aging **7**, 546–550.
Turner, C. W., Smith, S. J., Aldridge, P. L., and Stewart, S. L. (**1997**). "Formant transition duration and speech recognition in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **101**, 2822–2825.
Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). "Speaking clearly for the hard of hearing: IV. Further studies of the role of speaking rate," J. Speech Hear. Res. **39**, 494–509.
Vaughan, N. E., and Letowski, T. (**1997**). "Effects of age, speech rate, and type of test on temporal auditory processing," J. Speech Lang. Hear. Res. **40**, 1192–1200.
Wechsler Adult Intelligence Test (**2004**). 3rd ed., Dutch version (Harcourt Test, Amsterdam).
Willott, J. F. (**1991**). *Aging and the Auditory System: Anatomy, Physiology and Psychophysics* (Thompson, San Diego).
Wingfield, A., Tun, P. A., Koh, C. K., and Rosen, M. J. (**1999**). "Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech," Psychol. Aging **14**, 380–389.
Witton, C., Stein, J. F., Stoodley, C. J., Rosner, B. S., and Talcott, J. B. (**2002**). "Separate influences of acoustic AM and FM sensitivity on the phonological decoding skills of impaired and normal readers," J. Cogn. Neurosci. **14**, 866–874.
Witton, C., Talcott, J. B., Hansen, P. C., Richardson, A. J., Griffiths, T. D., Rees, A., Stein, J. F., and Green, G. G. R. (**1998**). "Sensitivity to dynamic auditory and visual stimuli predicts nonword reading ability in both dyslexic and normal readers," Curr. Biol. **8**, 791–797.
Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., and Merzenich, M. M. (**1997**). "Deficits in auditory temporal and spectral resolution in language-impaired children," Nature (London) **387**, 176–178.
Zekveld, A. A., George, E. L. J., Kramer, S. E., Goverts, S. T., and Houtgast, T. (**2007**). "The development of the text reception threshold test: A visual analogue of the speech reception threshold test," J. Speech Lang. Hear. Res. **50**, 576–584.

A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data

Richard E. Turner[a)]
Gatsby Computational Neuroscience Unit, Alexandra House, 17 Queen Square, London WC1N 3AR, United Kingdom

Thomas C. Walters,[b)] Jessica J. M. Monaghan,[c)] and Roy D. Patterson[d)]
Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

(Received 17 April 2007; revised 12 January 2009; accepted 12 January 2009)

This paper investigates the theoretical basis for estimating vocal-tract length (VTL) from the formant frequencies of vowel sounds. A statistical inference model was developed to characterize the relationship between vowel type and VTL, on the one hand, and formant frequency and vocal cavity size, on the other. The model was applied to two well known developmental studies of formant frequency. The results show that VTL is the major source of variability after vowel type and that the contribution due to other factors like developmental changes in oral-pharyngeal ratio is small relative to the residual measurement noise. The results suggest that speakers adjust the shape of the vocal tract as they grow to maintain a specific pattern of formant frequencies for individual vowels. This formant-pattern hypothesis motivates development of a statistical-inference model for estimating VTL from formant-frequency data. The technique is illustrated using a third developmental study of formant frequencies. The VTLs of the speakers are estimated and used to provide a more accurate description of the complicated relationship between VTL and glottal pulse rate as children mature into adults.
 [DOI: 10.1121/1.3079772]

PACS number(s): 43.72.Ar [DOS] Pages: 2374–2386

I. INTRODUCTION

The purpose of this paper is to establish a framework for estimating the vocal-tract length (VTL) of a given speaker from small segments of their voiced speech sounds and thereby to establish a method for continuous estimation of VTL during speech processing. A VTL track, or contour, would assist speaker segregation and vowel normalization in multi-source environments (e.g., Welling and Ney, 2004). The principle is illustrated in Fig. 1 which shows the magnitude spectra (vertical lines) of two synthetic /i/ vowels like those that might be produced by (a) a small child (about 95 cm in height) with a short vocal tract (9.4 cm) and (b) a tall woman (about 188 cm in height) with a long vocal tract (15 cm). The glottal-pulse rate (GPR) and the spectral envelope of a real vowel can be modified in an analogous way using the high quality vocoder referred to as "STRAIGHT" (Kawahara *et al.*, 1999; Kawahara and Irino, 2004). For simplicity, the GPR (the voice pitch) is 200 Hz in both cases; it determines the spacing of the spectral components. The spectral envelopes of the vowels are shown by the smooth lines; they represent the transfer functions of the vocal tracts that produced these vowels. The soft-shouldered peaks in the envelopes represent the vocal-tract resonances which are referred to as formants. The spectra and envelopes are plotted on a logarithmic frequency axis (base 2) and the reference frequency associated with a log value of 0 is 100 Hz.

It is assumed that the child and the adult have formed their vocal tracts into the same shape to produce their /i/ vowels, and as a result, the pattern of formant information is the same for the two tokens of the vowel, on this logarithmic frequency axis. It is only the position of the pattern that differs; the formant pattern for the adult is shifted toward the origin, with respect to that of the child, because the vocal tract of the adult is much longer than that of the child. This very simple model of vowel development is referred to as the fixed-formant-pattern hypothesis in the current paper. It is similar to the uniform scaling hypothesis (e.g., Fant, 1966) and to formant ratio theory (e.g., Miller, 1989). The differences will be discussed below where relevant.

These simple models of vowel development are important because they imply that vowel-type information and VTL information are *covariant* in the log-frequency vowel spectrum (Patterson *et al.*, 2007), and this, in turn, suggests a relatively simple method for performing VTL estimation in conjunction with vowel identification. The process involves taking the formant pattern for each vowel type, in turn, and shifting it back and forth along the frequency axis to find out which formant pattern (vowel type) leads to the best fit, and the position in which this best fit is achieved. The position specifies the VTL of the speaker to within a fixed constant. Accordingly, the purpose of the paper is to review the developmental data pertaining to the spectral envelopes of vowel

[a)]Author to whom correspondence should be addressed. Electronic mail: turner@gatsby.ucl.ac.uk
[b)]Electronic mail: tcw24@cam.ac.uk
[c)]Electronic mail: jessica@ihr.mrc.ac.uk
[d)]Electronic mail: rdpl@cam.ac.uk

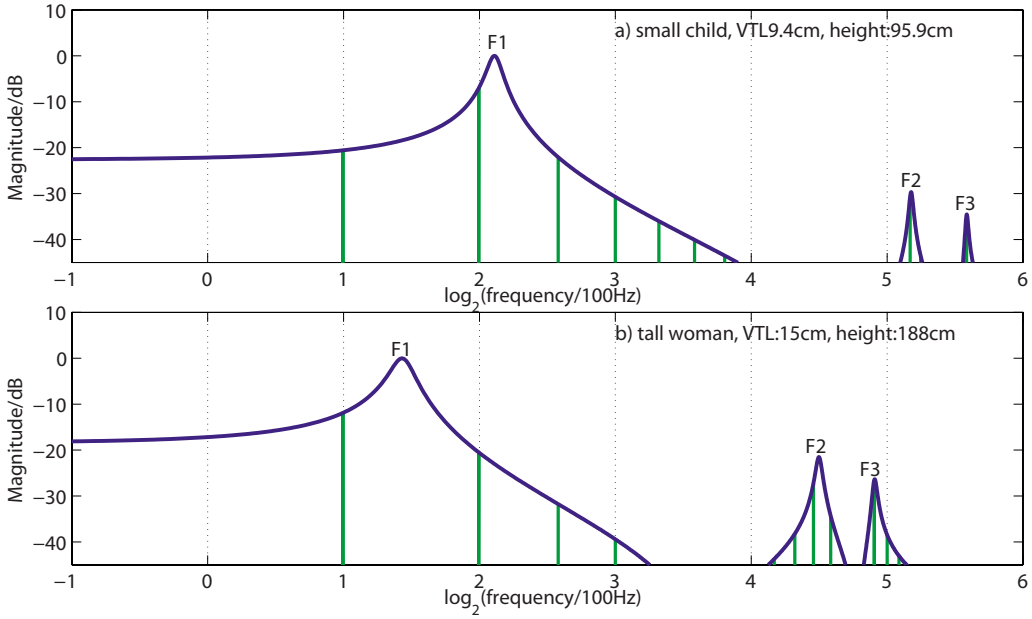


FIG. 1. (Color online) Magnitude spectra (vertical lines) and spectral envelopes (smooth lines) of two synthetic /i/ vowels like those that might be produced by (a) a small child and (b) a tall woman. F1, F2, and F3 show the positions of the first, second, and third formants of the vowel. Note that the formant peaks often occur between peaks of the magnitude spectrum.

sounds to determine the extent to which the formant-pattern hypothesis is true and to investigate the practical implications of estimating VTL from existing developmental data on formant patterns.

Section I A describes a quantitative reanalysis of the classic developmental data of Peterson and Barney (1952) in which the formant pattern is summarized by estimates of the first three formant frequencies, extracted from spectrograms of the vowels by humans. The analysis shows that children and adults produce vowels in which the formant pattern is, at least approximately, fixed, and the main source of variability, after vowel type, is the acoustic counterpart of VTL (i.e., acoustic scale; Cohen, 1993 and Umesh *et al.*, 1999). This suggests that it should be possible to use the formant-pattern principle to estimate VTL as outlined above from formant-frequency data. However, two problems were encountered when attempting to develop a procedure for VTL estimation. (1) There was a small but significant discrepancy from fixed formant patterns in the data of Peterson and Barney (1952) which suggested the presence of another source of variability in the data. Fant (1966, 1975) observed the discrepancy some time ago in the course of investigating the uniform scaling hypothesis. He suggested that the discrepancy might be associated with the fact that the pharyngeal cavity grows proportionately more than the oral cavity as children mature into adults. The extent of the non-uniformity of the growth of the anatomical cavities of the vocal tract is quantified in Sec. I B using the magnetic resonance imaging (MRI) data of Fitch and Giedd (1999). The analysis shows that the non-uniformity is actually much larger than would be required to explain the discrepancy from uniform scaling of formant frequencies in the data of Peterson and Barney (1952). (2) There is also an oddity in the Peterson and Barney (1952) data; there are an excessive number of formant frequencies which are integer multiples of the GPR of the vowel; that is, harmonics of the voice pitch. This suggests that the discrepancy might represent a measurement bias, arising from the well known problem of estimating formant frequencies from the harmonic spectra of voiced vowels. The problem is illustrated in the lower panel of Fig. 1. In this vowel token, there are harmonics (vertical lines) below and above the center of the first formant frequency (which is the first peak of the spectral envelope), but no harmonics right at the formant frequency. The problem is reviewed by de Cheveigné and Kawahara (1999) who concluded that it is inherent in spectrographic representations of speech sounds and extends to automatic transcription methods based on linear predictive coding (LPC).

The purpose of the remainder of the paper is to develop a model of vowel sounds that incorporates a statistical approach to deal with the formant measurement problem. This new approach to formant-frequency data has the distinct advantage that it can determine whether the residual formant variability in the data of Peterson and Barney (1952) is due to errors in formant-frequency estimation or to the non-uniform growth of one or more components of the vocal tract, as suggested by Fant (1975). The result is surprising; once the measurement noise has been properly modeled, it is observed that the formant patterns of the vowel sounds do not vary systematically, either with the size or the sex of the speaker, *despite the obvious non-uniformity* in the growth of the anatomical cavities (oral and pharyngeal). Moreover, a statistical analysis of the *fixed-formant-pattern* hypothesis at the heart of the model indicates that more complex growth functions with non-linear terms are less likely than the fixed pattern model, given the data of Peterson and Barney (1952). This means that the formant resonators are not affected by developmental changes in the oral-pharyngeal ratio and that it is reasonable to assume that formant-frequency values are effectively determined by vowel type and VTL, independent

of the position of the junction between the oral and pharyngeal cavities, as suggested by McGowan (2006). The results derived from the classic data of Peterson and Barney (1952) are confirmed by replicating the analyses on the massive developmental database reported by Lee *et al.* (1999). They recorded ten vowels from each of 436 children, ages 5–18, plus 56 adults. Section III shows how the statistical inference model can be used to estimate VTL from a separate set of developmental data reported by Huber *et al.* (1999) and to illustrate how GPR and VTL evolve from age 4 to adulthood in the human population.

A. Evaluating the fixed-formant-pattern assumption with the data of Peterson and Barney (1952)

The classic formant data of Peterson and Barney (1952) were reanalyzed to quantify the proportions of inter-vowel and intra-vowel variability and to assess the role of speaker size in the intra-vowel variability. Briefly, the analysis reveals that about 80% of the total variability in formant frequencies is accounted for by vowel type, and a second variable, which is referred to as acoustic scale (Cohen, 1993) and which is closely related to VTL and speaker height, accounts for up to 90% of the remaining intra-vowel variability. This indicates the potential value of VTL normalization for speech recognition as noted by, for example, Welling and Ney (2004), and it supports the hypothesis that it should be possible to estimate VTL from vowel sounds.

1. Inter-vowel variability: Vowel type

In their classic study, Peterson and Barney (1952) recorded two repetitions of ten American vowels in hVd words (heed /iy/, hid /ih/, head /eh/, had /ae/, hod /aa/, hawed /ao/, hood /uh, who'd /uw/, hud /ah/, and heard /er/) from 76 men, women, and children. From the spectrogram of each recording, they estimated the frequencies of the first three formants (F1, F2, and F3) and the pitch of the vowel (F0). When the data were plotted in F1-F2 space, the tokens of each vowel were found to cluster into relatively well defined regions that Peterson and Barney (1952) delimited with hand-drawn ellipses (their Fig. 8). In order to quantify the analysis, we have fitted three-dimensional Gaussian distributions to the F1-F2-F3 values of all of the tokens in each vowel cluster. The contours of constant probability associated with this distribution are ellipsoids; the contour associated with one standard deviation along each of the axes has been plotted for each of the ten vowels in Fig. 2. The formant frequency values have been converted into their corresponding wavelengths $(\lambda_1,\lambda_2,\lambda_3)$ because the focus of this paper is VTL and the analysis is more direct when presented in terms of wavelengths.

The positions of the ellipsoids in wavelength space reveal the established observations concerning inter-vowel variability. (1) There is virtually no overlap between the ellipsoids in this space. (2) The separation between the clusters is significantly greater in the λ_1-λ_2 plane than in the λ_2-λ_3 plane or the λ_1-λ_3 plane, indicating that the first two formants carry most of the vowel-type information. (3) The back vowels and front *unrounded* vowels occupy different

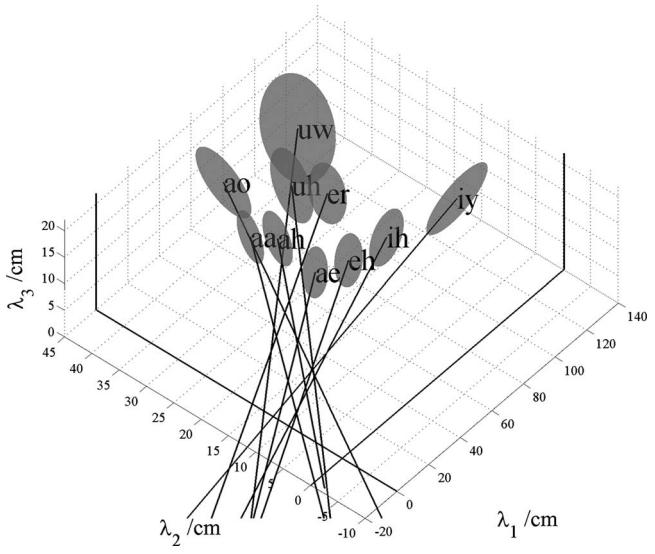


FIG. 2. Ellipsoids showing the distribution of formant wavelengths associated with each vowel in the classical formant data of Peterson and Barney (1952). The ellipsoids were derived using PCA; they represent one standard deviation from the mean of the vowel clusters (enclosing 29% of the data points). The lines show the orientation of the major axes of the ellipsoids; their extensions point toward the origin of the space, but there is a consistent bias away from the origin.

planes in wavelength space due to the relatively high and roughly constant second formant in front unrounded vowels (Broad and Wakita, 1977). The analysis shows that inter-vowel variability accounts for about 80% of the total formant variability in the Peterson and Barney (1952) data.

2. Intra-vowel variability: Vocal tract length

The intra-vowel variability is largely summarized by the eccentricity of the ellipsoid, its orientation, and its distance from the origin. With regard to *eccentricity*, in each case, one of the three principal axes of the ellipsoid is much longer than the other two. This is basically because the vocal tract increases in length as a child grows up. The eccentricity can be quantified with the aid of a principal components analysis (PCA), and it shows that approximately 90% of the intra-vowel variability lies in the direction of the major axis of the ellipsoid. With regard to the *orientation*, the ellipsoids all point in the direction of the origin of the space, as illustrated by the lines in Fig. 2; they show the extension of the major axis of each ellipsoid in the direction of the origin (given by the major-eigenvector of the covariance matrix). Together the eccentricities and orientations of the ellipsoids indicate that, within each vowel cluster, the formant pattern is approximately fixed for all members of the population. This is the basis of formant ratio theory (e.g., Lloyd, 1890; Potter and Steinberg, 1950; and Miller, 1989) and the uniform scaling hypothesis (e.g., Fant, 1966; for a review see Adank *et al.*, 2004). The principle has also been used to develop transforms intended to improve the performance of computer speech recognizers (e.g., Irino and Patterson, 2002; Umesh *et al.*, 2002; Welling and Ney, 2004). Similarly, it would appear that if vowel type and VTL account for virtually all of the variability in the formant-frequency data, then it would

TABLE I. Summary statistics for two tests of the fixed-formant-pattern hypothesis: The values in the upper pair of rows are based on a deterministic PCA; those in the lower pair of rows are based on a SFP model. Within each pair, the upper row shows the proportion of variability accounted for by the first component, and the lower row shows the angle that component makes with the uniform scaling line. Notice that the proportion of the variance explained by the principle component for the second analysis is often smaller than that for the first. This is because the second analysis explicitly models the observation noise and, once this contribution has been removed, there is less variability for the principal component to explain.

Vowel	/aa/	/ae/	/ah/	/ao/	/eh/	/er/	/ih/	/iy/	/uh/	/uw/	Ave
PCA: variability in direction of the PC	0.91	0.94	0.88	0.84	0.94	0.87	0.97	0.99	0.89	0.91	0.91
PCA: angle (deg)	11	14	6.0	7.5	8.5	13	6.6	5.6	3.9	8.0	8.4
SFP: variability in direction of the PC	0.85	0.95	0.97	0.85	0.96	0.70	0.96	0.87	0.98	0.83	0.89
SFP: angle (deg)	5.2	4.6	4.0	7.3	2.6	5.8	1.8	8.2	6.8	6.3	5.2

seem a relatively easy matter to estimate VTL given vowel type, as outlined in the Introduction (Sec. I).

3. Mathematical formulation of the formant-pattern model

The mathematical form for the simplest version of the formant-pattern model is

$$\underline{\lambda}_i^v = a_i \langle \underline{\lambda}^v \rangle, \tag{1}$$

where v is vowel type, i is the individual speaker, and $\underline{\lambda}_i^v$ is a three-component vector of formant wavelengths, which represents the formant pattern for vowel v of speaker i. $\langle \underline{\lambda}^v \rangle$ is the vector representation of the average formant pattern for vowel v in the population. The scalar, a_i, specifies the length of the individual's vocal tract relative to the mean of the population. The fixed-formant-pattern model is very simple; there is a single parameter for each formant and a single value of a_i that relates the individual's formants to those of the population. It is the "acoustic scale" (Cohen, 1993) of the formants relative to that of the population.

Peterson and Barney (1952) were not able to measure the VTL of their speakers; indeed, it is very difficult (Fitch and Giedd, 1999). However, the prediction is that variability in VTL across a population of speakers causes systematic variability in the formant wavelengths of their vowels. Specifically, the fixed-formant-pattern hypothesis predicts that the vowel clusters are ellipsoids and that the orientation of each vowel cluster, which is determined by the direction of the principal component of the variability in the data, will correspond to the direction along which the formant pattern is fixed. This direction, which is the major axis of the ellipsoid, passes through the origin. The eccentricity of the ellipsoid is partly determined by the variability of VTL in the population and partly by the distance of the ellipsoid from the origin of the space. In fact, the relative lengths of the ellipsoids along their major axes are predicted to depend entirely on their relative distances from the origin of the space according to

$$\sigma^v = \langle (\lambda^v - \langle \lambda^v \rangle)^2 \rangle^{1/2} = \sigma_a |\langle \underline{\lambda}^v \rangle|, \tag{2}$$

where σ^v is the magnitude of the principal component of the vowel cluster (which is equivalent to the length of the ellipsoid along the major axis), σ_a is the standard deviation of the VTL scalar in the population, and $|\langle \underline{\lambda}^v \rangle|$ is the magnitude of the vowel-cluster mean or the distance from the origin of the space to the center of the ellipsoid.

Thus, the model predicts that VTL is the largest source of intra-vowel variability and it can be used to assess the accuracy of the fixed-formant-pattern hypothesis. For example, the angle formed between the major axis of each ellipsoid and the line from the center of that ellipsoid to the origin provides a measure of the accuracy of the hypothesis. The angles are presented in Table I for the ten vowels in the Peterson and Barney (1952) database, along with the proportion of the intra-vowel variance accounted for by the principal component. The angles are very small and they show that VTL accounts for about 90% of the variability not attributable to vowel-type. Here, then, is a quantitative basis for the fixed-formant-pattern hypothesis as observed in the classic data of Peterson and Barney (1952).

4. Residual variability

Although the analysis of formant variability indicates that the fixed-formant-pattern hypothesis is largely correct, a detailed examination of Fig. 2 shows that when the main axes of the ellipsoids are extended toward the origin, they actually intercept the $\lambda_1=0$ plane at points where λ_2 and λ_3 are slightly, but consistently, positive, and they intersect the $\lambda_3=0$ plane at points where the values of λ_1 are consistently negative. This consistent bias in the intercepts suggests that there might be one more factor making a small, but consistent contribution to formant frequency variability. A clue to the form of the remaining variability is provided in Fig. 3 which shows the sub-clusters for men, women, and children plotted separately for six of the vowels in Fig. 2; the centroids of the sub-clusters for men, women, and children are relatively widely separated on the uniform scaling lines. This indicates that VTL variability is greater between speaker groups than within speaker groups—an observation that has been confirmed by Gonzalez (2004). Moreover, the principal axes of the sub-clusters with the more extreme values of λ_1 are more closely aligned with the first-formant axis than with the fixed formant-pattern line; the most obvious examples are the vowels /iy/ and /ih/. This suggests that within speaker sub-clusters, there may be another consistent source of variability *which is only revealed in conditions where VTL variability is small*.

There are several candidates for the source of this effect: Fant (1966, 1975) suggested that variability in the formant

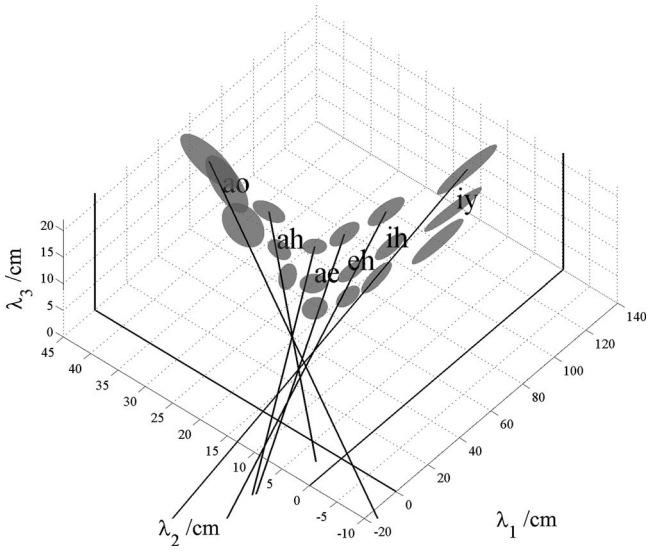


FIG. 3. Ellipsoids showing the distribution of formant wavelengths associated with the population subgroups (men, women, and children) for six vowels in the classical formant data of Peterson and Barney (1952). The ellipsoids were derived using PCA; they represent one standard deviation from the mean of the vowel clusters (enclosing 29% of the data points). The lines show the orientation of the major axes of the ellipsoids shown in Fig. 2; the sub-clusters of each vowel are well separated along the scaling line, but the principal axes of these sub-clusters are more closely aligned with the first-formant axis than the scaling line, especially for the vowels with a low first formant (e.g., /iy/).

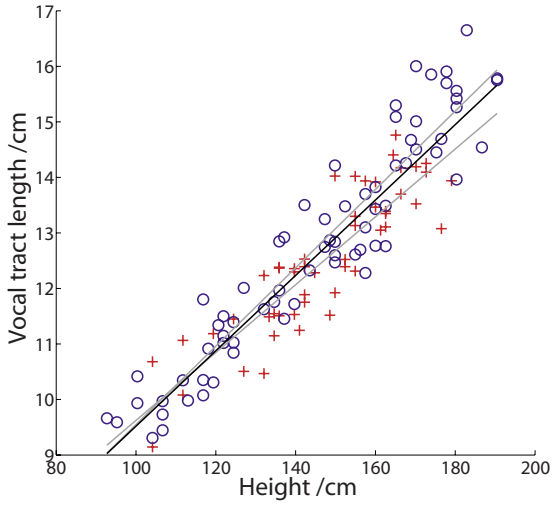


FIG. 4. (Color online) VTL as a function of height from the data of Fitch and Giedd (1999). The upper and lower faint trend lines are the fits to the males (circles) and females (crosses), respectively. The solid trend line is the fit to the male and female data combined.

pattern across speakers arises, at least in part, because the pharynx is proportionately larger in men than in women and children. He proposed a non-uniform scaling procedure with separate scale factors for each formant of each vowel to represent the non-uniform growth of the different components of the vocal tract and to take into account the changing formant-cavity-affiliations over vowels. The MRI data of Fitch and Giedd (1999) (reanalyzed below) confirm that the pharynx is proportionately larger in men, but this does not immediately indicate how the scale factors would be affected by VTL. Subsequently, Umesh *et al.* (2002) showed that Fant's (1966, 1975) scale factors could be averaged across vowels to form a single non-uniform scaling function that describes the scale factor as a function of formant frequency. In both cases, the implication is that there is one main, latent variable in this system, which is the acoustic scale of the vowel, but that this variable affects different formants in different ways, necessitating extra parameters to be added to the fixed pattern model. In Sec. I B, we develop a statistical model of formant-frequency data that can accommodate more complicated growth dependencies, should they be required, and which has the power to reveal any remaining sources of variability beyond those accounted for by vowel type and the fixed pattern hypothesis. In the event, however, what appears as an "effect" is revealed to be a bias caused by well known problems in formant-frequency estimation (e.g., de Cheveigné and Kawahara, 1999).

B. The form of the non-uniform growth of the oral and pharyngeal cavities

Fitch and Giedd (1999) used MRI to examine the growth of the components of the vocal tract as children mature into adults. The study included 129 men, women, and children ranging in age from 2.8 to 25 years. They recorded each subject's age, height, and weight, but they did not record samples of their speech sounds. The measurements were made with the subjects in the nasal breathing posture, and care was taken to exclude those who were overweight, or whose families had a history of language or developmental problems. Figure 4 shows VTL as a function of height separately for all of the males (○) (men and boys) and all of the females (+) (women and girls); VTL is effectively a linear function of height in both cases. There are proportionately more men at the tallest heights, but the two populations fall along lines with very similar slopes. It is also the case that the vocal tract grows proportionately slower than height because the head is proportionately larger than the body in children, but the proportionality is the same for the two groups. The growth rate is 0.067 cm/cm.

Figure 5 shows the relative lengths of the oral and pharyngeal portions of the vocal tract as a function of VTL, separately for males (○) and females (+). The figure shows that the length of the oral cavity decreases, and the length of the pharyngeal cavity increases, *relative to* VTL, as VTL increases. This is because the size of the oral cavity is largely determined by the size of the head which decreases as a proportion of body height as a person grows up. The figure makes it clear that the growth of the oral and pharyngeal cavities is *decidedly* non-uniform. Note, however, that the changes are linear in these coordinates and, for a given VTL, there is no difference between males and females in terms of the *proportions* of the cavities. This suggests that models, which relate growth in the main vocal tract cavities to the progressive decrease in formant frequencies with age, need not be excessively complex.

It is the pronounced non-uniform growth of the oral and pharyngeal cavities that prompts us to avoid the phrase "uniform scaling" when describing formant-frequency variability and to adopt instead the phrase "fixed formant pattern." The phrase uniform scaling is too readily misinterpreted as im-

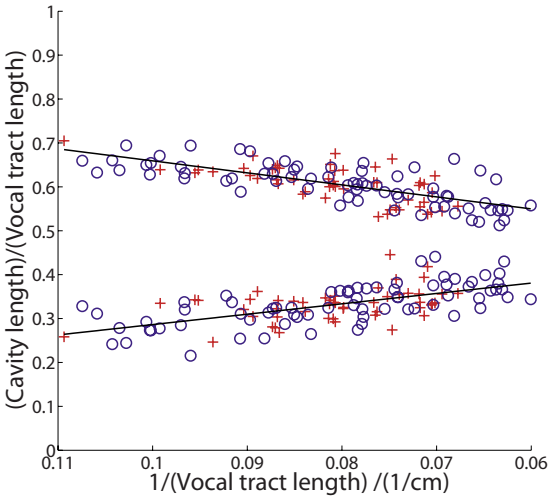


FIG. 5. (Color online) Growth functions for the oral cavity (upper cluster) and the pharyngeal cavity (lower cluster) from the data of Fitch and Giedd (1999). The abscissa is the reciprocal of VTL which orders the subjects according to size from left to right across the cluster. The ordinate is the relative length of the cavity and so the figure illustrates the change in the oral-pharyngeal ratio as people grow up. The trend lines show linear fits to the two clusters. The data of the males (circles) and females (crosses) cluster along the same line in both cases.

plying that the consistency of the formant pattern is the result of uniform growth of the anatomical components of the vocal tract, which is clearly incorrect in the case of the oral and pharyngeal cavities.

The fact that the formant patterns of vowel sounds do not vary markedly, either with the size of the speaker or their sex, may in retrospect seem surprising, given the striking non-uniformity in vocal-tract growth illustrated in Fig. 5. The analysis suggests that the anatomical distinction between the oral and pharyngeal divisions of the vocal tract is immaterial to the acoustic result of speech production. For a given vowel, the tongue constriction is simply positioned where it produces the appropriate ratio of front-cavity length to back-cavity length, independent of the location of the oral-pharyngeal junction

II. A STATISTICAL VERSION OF THE FORMANT PATTERN MODEL

The fact that the growth functions for the oral and pharyngeal cavities are linear suggests that it might be fairly simple to incorporate these growth functions into a model of formant-frequency data and in so doing realize Fant's (1966, 1975) original ambition. It is necessary, however, to adopt statistical methods to learn the relationship between formant frequency and acoustic scale from the data, and it is necessary to treat the acoustic scale of the vowel as a "latent variable" which is to be inferred from the data. Broadly speaking, the purpose of this section is to construct a statistical model of formant-wavelength data, which can determine the extent to which the complexities in the formant data of Peterson and Barney (1952) are due to non-uniform vocal tract growth, which needs to be included in the model, and the extent to which the complexities are attributable to measurement noise and are therefore artefactual. The development of this statistical formant pattern (SFP) model begins in Sec. II A, where a linear-model is developed to represent the growth functions of the oral and pharyngeal cavities, and the model is shown to be consistent with the developmental data of Fitch and Giedd (1999). Section II B explains why the model has to include explicit terms for the variability of the individual formants and why acoustic scale has to be incorporated as a latent variable. Section II C explains that it is also necessary to include explicit terms for the variability of the components of the measurement noise. Finally, in Sec. II D, the SFP model is applied to the developmental data of Peterson and Barney (1952), and then to the developmental data of Lee *et al.* (1999), to determine whether non-uniform vocal-tract-growth factors need to be incorporated into formant data models.

A. Modeling vocal tract length

As children mature and their height increases, so does the length of their vocal tract and thus the acoustic scale of their speech sounds. Height, VTL, and acoustic scale also depend on the sex of the speaker beyond about age 12 when VTL and acoustic scale become somewhat greater in males relative to their height. The SFP model employs a latent or hidden variable to represent the general growth factor associated with developmental changes in formant frequencies; that variable is designated a. This general growth factor is directly related to both VTL and to acoustic scale (Cohen, 1993), and it is assumed to be multi-modal in the population with clusters corresponding to men, women, and children.

More specifically, it is assumed that the lengths of the various cavities and components of the vocal tract are linearly related to VTL and therefore to the growth factor, a, via the average length of the cavity, or component, and a weighting factor, which can be thought of as reflecting the growth rate of the cavity or component [see Eq. (3) below]. For cavities like the pharynx, where the proportion changes with growth, the dependence is strong and the weighting factor is large. For components like the lips, where the proportion changes little with growth, the dependence is weak and the weighting factor is small. The weighting factors enable us to construct a model of the vocal tract in terms of VTL, where the growth of the components of the vocal tract is non-uniform but, nevertheless, a *linear* function of height. Mathematically, the model is

$$L_c^v = \langle L_c^v \rangle + a\frac{dL_c^v}{da}, \tag{3}$$

where $\langle L_c^v \rangle$ is the average length for cavity c for people articulating the vowel v, and a is the relative VTL of the individual. dL_c^v/da is a constant that does not depend on the individual. This relationship is consistent with the analysis of vocal-tract component lengths presented by Fitch and Giedd (1999) as will now be shown. If the total length of the vocal tract is written as the sum of the component lengths as follows;

$$L^v = \sum_c L_c^v = \sum_c \left[\langle L_c^v \rangle + a \frac{dL_c^v}{da} \right] = \langle L^v \rangle + a \frac{dL^v}{da}, \tag{4}$$

this expression can be substituted back into Eq. (3) to eliminate the unknown acoustic scale factor, a, and produce an expression for "the ratio of a cavity or component's length," L_c, to "the total length of the vocal tract," L.

$$\frac{L_c^v}{L^v} = \left[\langle L_c^v \rangle - \langle L^v \rangle \frac{dL_c^v}{dL^v} \right] \frac{1}{L^v} + \frac{dL_c^v}{dL^v}. \tag{5}$$

Thus, in this model of vowel production, the growth of the individual cavities and components is predicted to be linear when plotted against the reciprocal of L, which is precisely what was observed in Fig. 3. There is variability in the data that the model does not absorb, but there does not appear to be any consistent deviation as a function of speaker height of the sort that would warrant including quadratic or higher-order terms in the model.

It is now possible to determine *quantitatively* whether higher-order terms are warranted by fitting Nth-order polynomials to the data, and learning maximum-likelihood parameters for the terms and corresponding error-bars on these inferences. The linear model can then be compared to models of higher order by weighting the best-fit likelihoods of the more complicated models by penalty factors known as Occam factors, which depend both on prior knowledge and the error bars on the maximum-likelihood parameter estimates. In Bayesian statistics, this is a non-arbitrary form of hypothesis test (Mackay, 2003). In the current case, the linear model is found to be much more probable than models with higher-order terms. Moreover, the linear terms in the higher-order approximations were found to have similar values to those of the linear model, and the higher terms were found to contribute little *within the range of the data*. This, then, is a quantitative justification for using the non-uniform but *linear* model of VTL variability, and it can now be used to deconvolve the effect of vocal tract changes on vowel formant frequencies.

B. The formant pattern model and the non-uniform growth of the oral and pharyngeal cavities of the vocal tract

The next step in the development of the model is to relate VTL to formant wavelength. Broadly speaking, the higher formants in Peterson and Barney's (1952) data, F2 and F3, are well modeled as simple standing wave resonances, so they will have wavelengths which are linearly dependent on the length of the vocal tract for a given vowel and formant. A simple standing wave is not, however, a good model of the first formant. *To wit*, the wavelength of the first formant can be as much as eight times the length of the vocal tract, which is twice the maximum length that would be expected for a simple standing wave resonance. Fant (1966) argued that the first formant is commonly a Helmholtz resonance, in which case, the relationship between the frequency of the first formant and the growth of the vocal tract might be expected to be more complicated. In the event, however, when the correlations between formants for all of the vowels were analyzed (using the Bayesian techniques described in Sec. II A) the relationships were found to be well approximated by a line. As a consequence, any pair of formants in a vowel is linearly related; that is,

$$\lambda_l^v = m_{l,m}^v \lambda_m^v + c_m^v, \tag{6}$$

where l and m are formant numbers, 1, 2, or 3. This means that a fairly simple model might be expected to capture the majority of the variability in Peterson and Barney's (1952) data, so long as it incorporates the model of vocal tract growth derived earlier in Sec. II A. A straightforward approach, consistent with the data, is to describe each resonator in terms of an effective wavelength, that is, a simple linear function of VTL, regardless of its physical complexity. That is, $\lambda_l^v = n_l^v L_l^v$. Each of the effective wavelengths might be expected to develop in exactly the same way as the physical dimensions of the vocal tract [Eq. (3)], in which case the predicted relationship between formant wavelengths is linear, as observed previously in this section. This description can be generalized to the three-component vowel vectors

$$\underline{\lambda}^v = \underline{c}^v + a\underline{m}^v \quad \text{where } \underline{c}^v = \underline{n}^v \circ \langle \underline{L}^v \rangle \text{ and } \underline{m}^v = \underline{n}^v \circ \frac{d\langle \underline{L}^v \rangle}{da}, \tag{7}$$

where $\circ$ denotes the element-wise product. The prediction of this model is that the vowel clusters will form on segments of lines oriented in the direction m^v with centroids at c^v. If the growth rate of the effective lengths is uniform, then m^v and c^v are parallel and the fixed-formant-pattern model is recovered as a simple limit. If the distribution of the acoustic scale factor, a, is Gaussian then this model is equivalent to PCA and the analysis of Sec. I A is recovered. However, as noted earlier, the distribution is not Gaussian; there are three distinct classes of speaker (men, women, and children). Therefore, a more sensible choice is a mixture of Gaussians, with a Gaussian component for each group.

Two versions of the statistical model were developed, distinguished by their assumptions concerning the source of the vowels in each vowel cluster. In the first and simpler version, the vowels in each cluster were treated as if they all came from different speakers, and thus the clusters can be fitted individually. In point of fact, the vowels in the clusters are not independent with respect to VTL; each speaker contributes two tokens to each of the ten vowel clusters. The second version of the model incorporates this constraint, which, in turn, makes it possible to fit all the vowel clusters simultaneously. Although the inferred acoustic scale factors estimated with the second version of the statistical formant pattern model are almost certainly more accurate, the parameter values derived from the two models are very similar. Accordingly the discussion is restricted to the results from the second version of the model, and it is these values that are reported in the lower rows of Table I.

C. The variability of formant measurements

Having included the effects of vowel type and VTL in the statistical version of the formant pattern model, the question is whether the formant-frequency data contain other

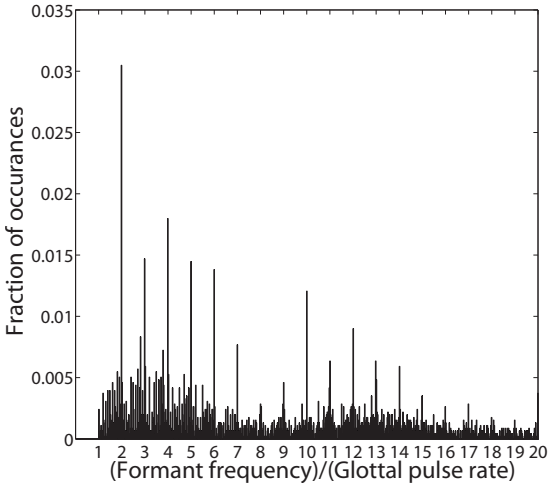


FIG. 6. Histogram of formant frequencies normalized to glottal pulse rate, illustrating the over-abundance of formant frequencies which are integer multiples of the GPR (i.e., GPR harmonics).

consistent sources of variability or whether the remaining variability is just due to measurement noise. To answer the question, it is necessary to include an explicit term for the residual noise in the SFP model. When the model, with its noise term, is applied to the data, the result is surprising; most of the remaining variability in the formant wavelength data is due to a consistent measurement error, and when the error is properly modeled, the fixed-formant-pattern model is observed to absorb most of the remaining variability. This indicates that if there is another natural factor, then its effect is limited to a very small contribution—a contribution that would be difficult to characterize because its effect is obscured by measurement noise.

The measurement error arises from the fact that it is difficult to estimate formant-frequency values from a spectrogram, particularly for the first formant, as noted in the Introduction and illustrated in Fig. 1. Formant-frequency estimation based on LPC is also prone to errors in this situation; it can only guarantee accuracy of approximately a quarter of the glottal pulse rate (Monsen and Engebretson, 1983; Vallabha and Tuller, 2002). Peterson and Barney's (1952) method was less sophisticated; they used a simple weighted average of the harmonics, f_n, in the neighborhood of the formant (see Potter and Steinberg, 1950),

$$\text{formant frequency} = \frac{\Sigma_n f_n \times a_n}{\Sigma_n a_n}. \tag{8}$$

Statistically, their method has similar restrictions to those of LPC with respect to accuracy, but the observed errors are somewhat larger. Moreover, an analysis of the data shows that a curiously high proportion of the formant estimates ($\sim$20%) are integer multiples of the GPR, as shown in Fig. 6. It appears that the estimate of formant frequency is attracted by a nearby harmonic frequency, which it would only rarely be by chance. It is also clear that many of the formant-frequency estimates were based on a single harmonic frequency. It turns out that this consistent measurement error is the source of much of the remaining variability in the vowel formant data. The measurement noise is roughly the same in

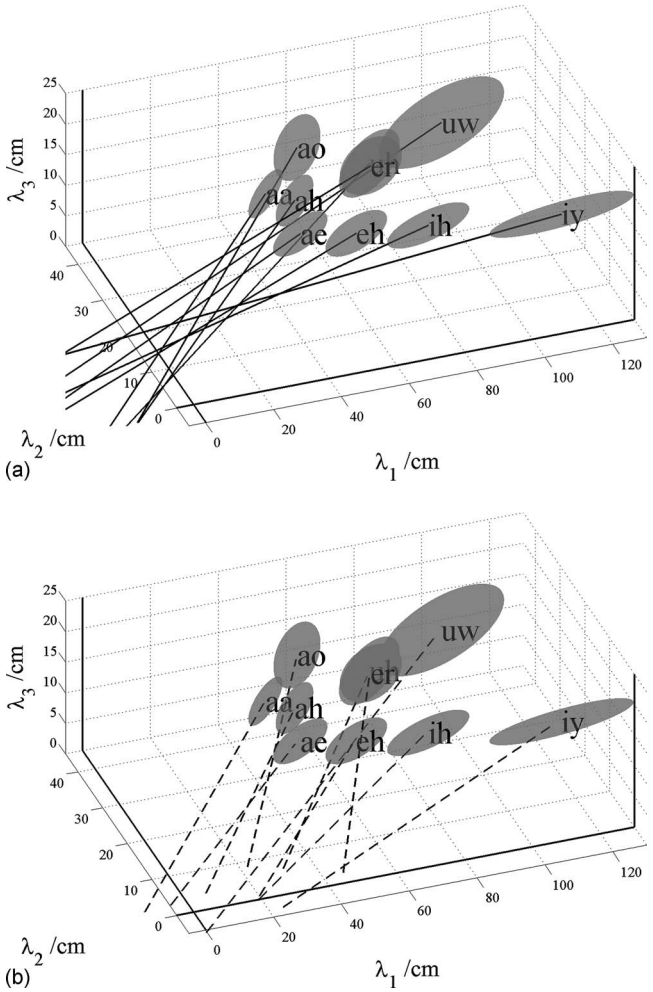


FIG. 7. (a) Rotated view of Fig. 2, showing the composite ellipsoids for the vowels in Peterson and Barney (1952), to emphasize the bias of the intercepts away from the origin in the λ_1 dimension. (b) The same view of the composite ellipsoids using the statistical formant-pattern model with explicit terms for formant measurement error. The bias of the intercepts is considerably reduced.

absolute terms for all of the formants and so, as a proportion, the effect is largest for the first formant and smallest for the third formant. In wavelength terms,

$$\frac{\sigma_\lambda}{\lambda} = n\frac{f_0}{f}. \tag{9}$$

The distortion that the measurement error imparts to the vowel clusters is illustrated in Figs. 7(a) and 8(a), which show views of the respective spaces presented in Figs. 2 and 3, but with the view rotated to emphasize the λ_1-λ_3 plane and thereby to emphasize the variability in λ_1. In Fig. 8(a), for vowels having a large λ_1, such as /iy/ and /ih/, the individual ellipsoids for men, women, and children are observed to be highly elongated in the λ_1 direction. The elongation is a very unusual form of variability, and it is not clear how factors like the non-linear growth of the oral-pharyngeal ratio could explain this form of variability since it does not vary with speaker size. The form of the variability led to the hypothesis that the lack of an explicit noise term in the deterministic formant pattern model (Sec. I A) leads to a bias in formant wavelength estimation that, in turn, produces the elongation

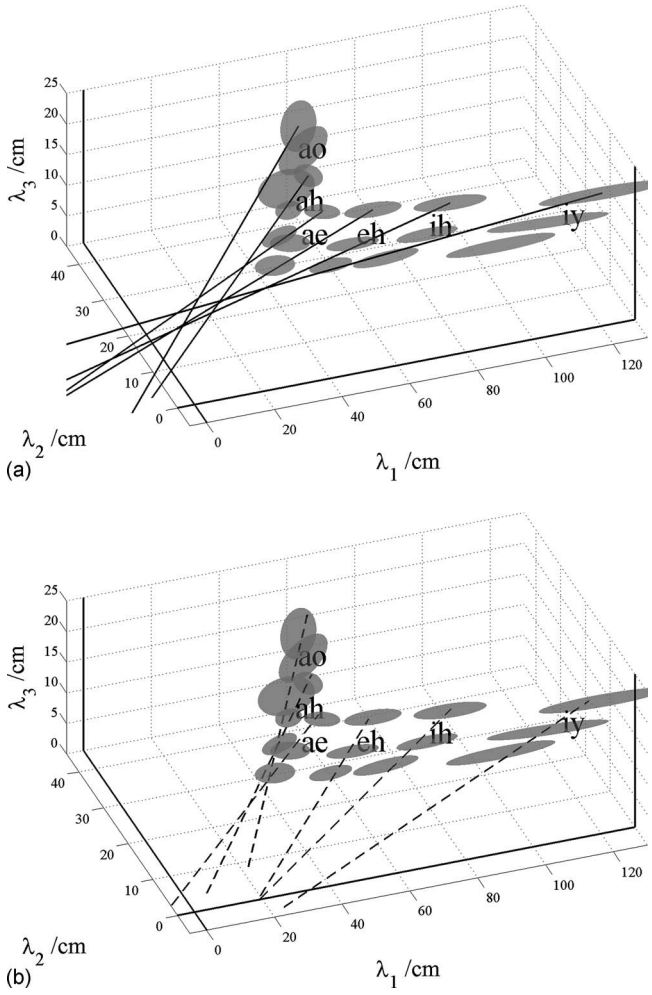


FIG. 8. (a) Rotated view of Fig. 3, showing the individual ellipsoids for the population subgroups for six of the vowels in Peterson and Barney (1952). The view emphasizes the bias of the intercepts away from the origin in the λ_1 dimension, and the elongation of the ellipsoids in the λ_1 dimension for vowels /iy/ and /ih/. (b) The same view of the individual ellipsoids using the statistical formant-pattern model with explicit terms for formant measurement error. The bias of the intercepts is considerably reduced.

of the ellipsoids in Fig. 8(a). It also produces the twisting of the angle of the main axis of the composite ellipsoid, as observed in Figs. 2 and 7(a), which, in turn, causes the intercept of the axis to shift away from the origin. The generation of the bias is illustrated schematically in Fig. 9; Sec. II D shows how to model the noise statistically to eliminate the bias in the full SFP model. The SFP model avoids the bias by introducing explicit terms for the noise associated with each formant,

$$\underline{\lambda}^v = \langle \underline{\lambda} \rangle^v + a\underline{m}^v + \underline{\eta}^v + \sum_{j=1}^{J} b_j^v \underline{n}_j^v, \tag{10}$$

where $\underline{\eta}^v$ is the formant-specific noise term. It is a vector of zero mean Gaussian noise with covariance given by $\underline{\underline{\Psi}}^v = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. A factor, b_j^v, was also added to capture any other consistent source of natural variability in the data.

This model of formant wavelengths is a modified version of factor-analysis (FA) (Roweis and Ghahramani, 1999), where the distribution over the latent variable, $p(a)$, is a mixture of Gaussians (one for each speaker type; man, woman, and child) rather than a single Gaussian. The mixture of Gaussians was used to represent the statistics of the noise, $p(b)$, as well, but the divergence from a simple Gaussian was found to be minimal.

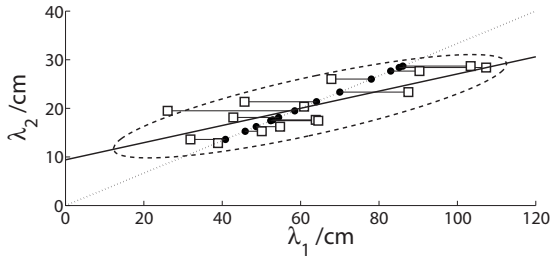


FIG. 9. Illustration of how measurement noise can bias the orientation of the major axis of the ellipsoid derived with a traditional deterministic analysis of noisy data. The true data (circles) lie on a uniform scaling line (dotted line), but are corrupted by anisotropic measurement noise (squares). Both the values of the measurement noise and the formant frequencies have been chosen to be typical of the vowel /iy/. A traditional fit with PCA derives a biased ellipse (dotted) with a principal axis which is biased away from the uniform scaling line. The more sophisticated analysis developed in this paper, which can learn the differing contributions from observation noise in each dimension, can recover the true directions, and it is unbiased in this regard.

In order to assess the relative contributions of an extra natural factor, on the one hand, and measurement noise, on the other, and to infer the acoustic scale factor of an unknown speaker from their formant data, the values of the model's parameters must be learned from the data. That is, they must be learned from the statistics of (i) the relative vocal tract lengths, $p(a)$, (ii) the noise, $\underline{\underline{\Psi}}^v$, and (iii) the factor loadings, $\underline{m}^v$. The learning and inference is accomplished using probabilistic methods, in particular, the variational expectation-maximization (EM) algorithm of Ghahramani and Hinton (1996). This algorithm repeatedly optimizes a lower bound on the log-likelihood of the parameters in two steps: in the expectation (E) step, the algorithm infers the acoustic scale factors of the speakers, given the current parameter estimates; in the maximization (M) step it finds the most likely parameters given the inferred acoustic scale factors. The iteration of the two steps typically converges in the region of the maximum-likelihood estimate for the parameters (Ghahramani and Hinton, 1996).

D. Application of the statistical, formant-pattern model to developmental formant data

We are now in a position to assess the relative contribution from the hypothetical extra natural factor, $\underline{n}_j^v$, on the one hand, and measurement noise, $\underline{\underline{\Psi}}^v$, on the other hand. The result is illustrated for the Peterson and Barney (1952) data in Figs. 7(b) and 8(b); they showed the major axes of variability for the vowel ellipsoids after they have been corrected for measurement noise. These corrected vectors point more accurately toward the origin of the space, as would be predicted by the fixed-formant-pattern hypothesis, indicating that there is no need for an extra natural factor to explain this data set (that is, $\underline{n}_j^v = 0$). The orientation of each component derived with the statistical model was compared to the orientation of the corresponding component derived from the original analysis. The results are presented in the lower rows

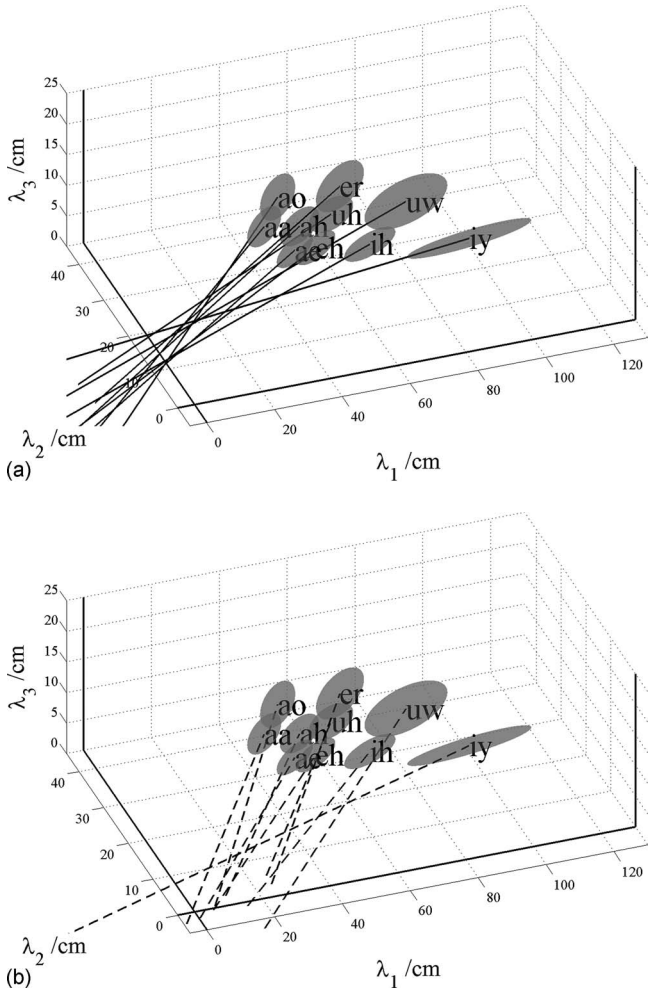


FIG. 10. (a) The composite ellipsoids for the vowels in Lee *et al.* (1999). The view emphasizes the bias of the intercepts away from the origin in the λ_1 dimension. (b) The same view of the composite ellipsoids using the statistical formant-pattern model with explicit terms for formant measurement error. The bias of the intercepts is considerably reduced.

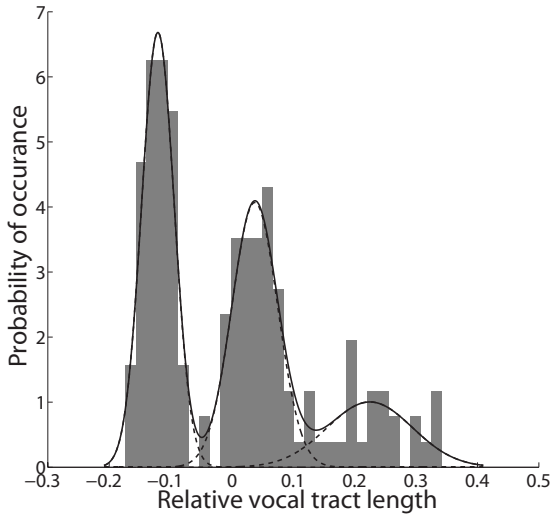


FIG. 11. Histogram of VTLs derived from the data of Peterson and Barney (1952) using the statistical formant-pattern model. The mixture of Gaussians derived by the model is shown by the solid line; the Gaussians which make up the fit are shown by dotted lines.

of Table I; they show that if measurement noise and other sources of natural variability are modeled statistically, the component of variability attributable to VTL becomes more uniform, while the residual noise decreases correspondingly.

In more recent developmental studies of formant frequency, such as those of Hillenbrand *et al.* (1995) and Lee *et al.* (1999), the formant-frequency values were estimated automatically from spectral frames of vowel sounds using LPC. The deterministic and statistical versions of the formant-pattern model were fitted to the data of both Hillenbrand *et al.* (1995) and Lee *et al.* (1999) to determine (a) whether the extended axes of the vowel ellipsoids would still show the bias away from the origin when the deterministic version of the model was fitted to the LPC data and (b) whether the bias would be reduced when the statistical version was fitted to the data.

The results for the data of Lee *et al.* (1999) are presented in Fig. 10 in the same format as shown for the data of Peterson and Barney in Figs. 7 and 8. The database of Lee *et al.* (1999) is far larger than that of Hillenbrand *et al.* (1995) and it covers a much greater range of ages. Lee *et al.* (1999) recorded ten vowels spoken by 436 children, ages 5–18, and 56 adults. Figure 10(a) shows the location of the ellipsoids for each vowel in wavelength space. The distribution of the ellipsoids is very similar to that shown for the Peterson and Barney (1952) data in Fig. 7(a), and the extensions of the major axes show the same bias away from the origin. There is also the same pronounced elongation of the ellipsoid for the vowel /iy/ in the λ_1 direction, probably due to the problem of estimating λ_1 which is particularly long in this vowel. The results for the SFP model are presented in Fig. 10(b), where the main axes of the ellipsoids are observed to intercept the λ_1-λ_3 plane at points much closer to the origin, indicating that the average bias is considerably reduced for this version of the formant-pattern model. Similar results were obtained with the data of Hillenbrand *et al.* (1995) indicating that the LPC method of extracting formant frequencies has a similar problem to that observed with the spectrogram reading method, as would be expected. The figure for the data of Hillenbrand *et al.* (1995) is omitted for brevity.

Finally, the assumptions made in Sec. II B, concerning the distribution of acoustic scale factors and the observation noise, need to be verified. With regard to the distribution of acoustic scale factors, which was modeled as three Gaussian sub-populations for men, women, and children, Fig. 11 shows the inferred scales together with the fitted mixture of Gaussians. The distribution is clearly multi-modal, justifying the assumptions of the model. With regard to the measurement noise, the SFP model learns that the average error is 50 Hz which is consistent with the inherent inaccuracy in the formant extraction process described in Sec. II C.

The formant measurement error produces a particular problem for deterministic versions of formant ratio theory (e.g., Miller, 1989), inasmuch as ratios accentuate variability. Moreover, it is traditional to use ratios that have F1 in the denominator, and F1 is the formant estimate that is most prone to error, so the accentuation of the variability is particularly large in the traditional version of formant ratio theory. The variability introduced by the measurement error and the accentuation of the variability associated with the use

of ratios are likely to have hampered efforts to normalize for acoustic scale using formant ratios. In the statistical analysis of formant-frequency data, the vectors of formant frequencies are treated as stochastic patterns and the three formants are fitted simultaneously, without the calculation of ratios. This shift in emphasis is important and it is why the model is referred to as a statistical formant-pattern model rather than formant ratio theory. The phrase "formant pattern" is intended to emphasize that the vector of formant wavelengths is a representation of the spectral envelope of the vowel. The shape and position of the spectral envelope are best estimated using a vector of formant peaks which are statistically defined, which is why the acronym for the model includes the "S." To reiterate, deterministic formant ratios provide a rather unreliable measure of the spectral envelope of a vowel because they amplify the error of the formant in the denominator and, unfortunately, it is common practice to use the most error prone of the formants as the denominator.

III. ESTIMATING VTL FROM FORMANT-FREQUENCY DATA

In this section, the SFP model is extended to produce an algorithm for estimating VTL, which involves calibrating the acoustic scale factors to a measure of absolute size. The algorithm is then applied to the developmental formant data of Huber *et al.* (1999) to estimate the VTLs of their speaker groups. Huber *et al.* (1999) reported the average values for the first three formants in /a/, separately, for groups of ten males and ten females in each of nine age bins (4, 6, 8, 10, 12, 14, 16, 18, and adults). Developmental studies of formant frequency often include developmental data on GPR and a comparison of the growth rates for VTL and GPR, since the growth rates change radically at puberty. The study of Huber *et al.* (1999) includes GPR data which we combine with the VTL estimates to illustrate the complicated developmental relationship between the excitation and resonance components of vowel production.

A. Inferring a speaker's VTL: Calibration of the data of Huber *et al.* (1999) to the VTLs of Fitch and Giedd (1999)

The main issue in the VTL estimation algorithm, as in the vowel production model, is to identify and correctly characterize the different components of the variability in the formant-frequency measurements. It is assumed that the measurement noise for individual vowel samples in the data of Huber *et al.* (1999) has approximately the same form as that in the Peterson and Barney (1952) data. The articulation of the vowels is assumed to be the same in the two studies. Both of these assumptions can be verified retrospectively. The methods developed in Sec. II can now be used to infer the acoustic scale factors for each group of speakers in each age category. In order for these estimates to be converted into absolute VTLs, we need to assume that the studies of Huber *et al.* (1999) and Fitch and Giedd (1999) were sampling from the same distribution of people, which was modeled as a mixture of Gaussians earlier. In this case, the relative VTLs from the study of Huber *et al.* (1999) can be scaled using the mean and variance data from Fitch and Giedd (1999). The VTLs inferred from the study of Huber *et al.* (1999) and the lengths measured by Fitch and Giedd (1999) are presented together in Fig. 12. The correspondence is surprisingly good, particularly for the male speakers. The male values from the study of Huber *et al.* (1999) seem a little high at 4 and 6 years, and the female values seem a little low for age 16 and above, but the deviations are not large relative to the overall variability.

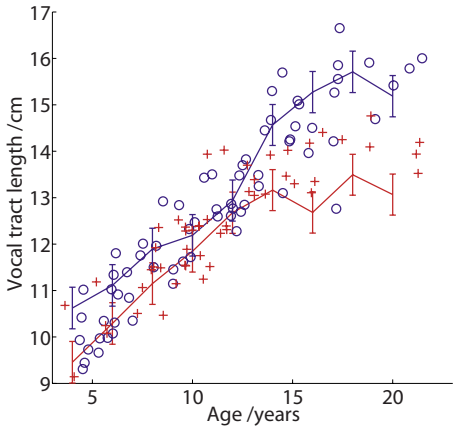


FIG. 12. (Color online) VTLs inferred for men and women from the data of Huber *et al.* (1999), plotted as a function of the speaker's age. For comparison, the data of Fitch and Giedd (1999) are presented by circles for men and crosses for women. The correspondence is impressive given that the populations in the two studies are not the same.

B. The GPR-VTL plane: Development and natural variability

Finally, the data from Peterson and Barney (1952) and Huber *et al.* (1999) were combined to characterize the developmental trajectory of vowel sounds in the log GPR-log VTL plane. The domains occupied by men, women, and children in the plane were delineated using the Peterson and Barney (1952) data. The 20 vowel sounds produced by each speaker (two tokens of each of ten vowels) were used to produce an estimate of each speaker's VTL, using the EM algorithm and the procedure described in Sec. III A; when combined with the corresponding GPR estimates, each individual provides 20 GPR-VTL points on the GPR-VTL plane. Two-dimensional Gaussian distributions were fitted separately to the data of the men, women, and children to characterize the domain of each speaker class on the plane. Contours of constant probability in these distributions are elliptical in form, and the contours that enclose about 80% of the individuals in each speaker class are shown by the three ellipses in Fig. 13. Unfortunately, the record of the Peterson and Barney (1952) data currently available does not contain information regarding the ages of individual children or their heights, and so the VTL estimates of the children had to be calibrated using the values for the adult males. Specifically, the mean and variance for the VTL estimates of the adult males [derived from the Peterson and Barney (1952) data] were equated to the mean and variance of the VTL values for the adult males reported by Fitch and Giedd (1999).

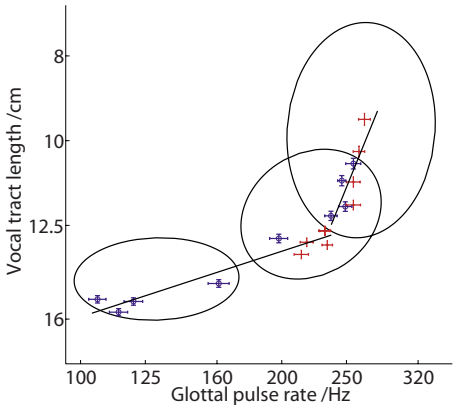


FIG. 13. (Color online) Predicting the population distribution over the log GPR-log VTL plane and the path of an average male and female through it as they develop. The distributions of inferred VTLs and GPRs for the data of Peterson and Barney (1952) are characterized by the three ellipses representing men (lower left), women (center), and children (upper right). The ellipses are drawn at two standard deviations about the mean to enclose ~80% of the data points. The developmental paths derived from the data of Huber *et al.* (1999) for the average female and the average male are shown by circles and crosses, respectively.

The mean VTL value inferred for each age-by-sex group in the data of Huber *et al.* (1999) was paired with the appropriate mean GPR value and plotted in the logGPR-logVTL plane of Fig. 13 to show the developmental trajectory of the voice for males and females as they mature. The symbols include ± one standard deviation in both dimensions. Within each ellipse, the trajectory from the data of Huber *et al.* (1999) reflects the eccentricity of the ellipse derived from the Peterson and Barney (1952) data. There is good agreement between the developmental data and the positions of the ellipses. It appears that the growth trajectories can be summarized with a pair of straight lines that meet near the center of the ellipse for women. The segment with the steeper slope was fitted to the data of males and females from ages 4 to 10; it has a slope of 1.9, so the VTL of a child is roughly proportional to the square of their GPR. The segment with the shallower slope was fitted to the data of males and females from age 12 upward; it has a slope of 0.25, so the VTL of adults and adolescents is roughly proportional to $GPR^{1/4}$. The figure makes it clear that the developmental trajectory for males changes dramatically around puberty and the change in trajectory is mainly due to the sudden drop in pitch that occurs at this time.

With regard to practical application of the VTL estimation process, it is theoretically possible to estimate VTL from all vowels and sonorant consonants on a frame by frame basis, *when the vowel type is known*. However, VTL would be expected to change at a much slower rate than formant frequency since it mainly changes with the speaker. So a reasonable strategy would probably be to limit VTL estimation to strong vowels where the recognizer is confident of the vowel type.

IV. SUMMARY AND CONCLUSIONS

A PCA was used to cluster the classical formant-frequency data of Peterson and Barney (1952) and provide ellipsoids showing the distribution of formant frequencies associated with each vowel and population subgroup. The analysis revealed that vowel type accounts for 80% of the variability in formant frequencies and 90% of the remaining variability is accounted for by VTL. Sufficient variability remained to support the hypothesis that there might be another consistent source of variability, such as developmental changes in oral-pharyngeal ratio. The MRI data of Fitch and Giedd (1999) were reanalyzed to evaluate this hypothesis, and the analysis confirmed that the growth of the oral and pharyngeal cavities is non-uniform, with the pharyngeal cavity growing faster than the oral cavity. However, the growth functions are linear. Moreover, the growth functions for men, women, and children are all the same. Despite the non-uniform growth of the anatomical cavities of the vocal tract, there is no commensurate non-linearity in the formant-pattern data. Indeed, for a given vowel, the formant pattern is essentially fixed on a log-frequency axis, shifting toward the origin without changing shape as VTL increases. This means that the systematic variability in formant-frequency data (at least the first three formants) is effectively divided between vowel type and VTL, and it suggests that speakers adjust the shape of the vocal tract as they grow to maintain a specific spectral pattern for each vowel type, independent of the relative size of the oral and pharyngeal cavities. The conclusion is important because it means that it should be a straightforward matter to estimate VTL from the voiced sounds in continuous speech.

A statistical formant-pattern model of formant-frequency data was developed with (a) a latent variable to absorb the variability of all size related factors, (b) non-uniform, but linear growth functions for the oral and pharyngeal cavities, and (c) separate measurement-noise terms for each of the formants. A modified version of factor analysis was developed to infer the acoustic scale factor of the vowels and, thus, the VTL of an unknown speaker, from the formant data of a given vowel type. The use of statistical methods to model the measurement noise revealed that the vast majority of the variability not attributable to vowel type is associated with VTL, and if there are any other natural sources of systematic variability their contribution is small with respect to the error in formant frequency estimation. The statistical version of the formant-pattern model was used to correct the biases of the deterministic version of the model, and the effect of the correction was illustrated with the formant-frequency data of both Peterson and Barney (1952) and Lee *et al.* (1999).

Finally, the SFP model was used to analyze the developmental data of Huber *et al.* (1999). The mean VTL was estimated for each age group, and the results were used to chart the development of VTL and GPR in children as they mature into adults.

ACKNOWLEDGMENTS

We would like to thank Dr. T. Fitch for kindly providing the individual data on the lengths of the parts of the vocal tract from their MRI data. R.E.T. was supported by the Gatsby Charitable Foundation through the writing of this paper. The research was supported by the UK Medical Re-

search Council (G0500221 and G9900369) and by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under Grant No. FA8655-05-1-3043.

Adank, P., Smits, R., and van Hout, R. (**2004**). "A comparison of vowel normalization procedures for language variation research," J. Acoust. Soc. Am. **116**, 3099–3107.

Broad, D. J., and Wakita, H. (**1977**). "Piecewise-planar representation of vowel formant frequencies," J. Acoust. Soc. Am. **62**, 1467–1473.

de Cheveigné, A., and Kawahara, H. (**1999**). "Missing-data model of vowel identification," J. Acoust. Soc. Am. **105**, 3497–3508.

Cohen, L. (**1993**). "The scale transform," IEEE Trans. Signal Process. **41**, 3275–3292.

Fant, G. (**1966**). "A note on vocal tract size factors and non-uniform F-pattern scalings," QPSR Report No. 4, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm.

Fant, G. (**1975**). "Non-uniform vowel normalization," Speech Transmission Laboratory, Royal Institute of Technology, Stockholm. QPSR Report No. 2–3.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Ghahramani, Z., and Hinton, G. E. (**1996**). "The EM algorithm for mixtures of factor analyzers," University of Toronto Technical Report No. CRG-TR-96-1, http://www.learning.eng.cam.ac.uk/zoubin/papers.html (Last viewed January, 2008).

González, J. (**2004**). "Formant frequencies and body size of speaker: A weak relationship in adult humans," J. Phonetics **32**, 277–287.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–4111.

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., and Johnson, K. (**1999**). "Formants of children, women, and men: The effects of vocal intensity variation," J. Acoust. Soc. Am. **106**, 1532–1542.

Irino, T., and Patterson, R. D. (**2002**). "Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilized wavelet-Mellin transform," Speech Commun. **36**, 181–203.

Kawahara, H., and Irino, T. (**2004**). "Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Image Analysis Academic, Norwell, MA), pp. 167–180.

Kawahara, H., Masuda-Kasuse, I., and de Cheveigne, A. (**1999**). "Restructuring speech representations using pitch-adaptive time-frequency smoothing and instantaneous-frequency-based F0 extraction: Possible role of repetitive structure in sounds," Speech Commun. **27**, 187–207.

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.

Lloyd, R. J. (**1890**). "Speech sounds: Their nature and causation (I)," Phonetische Studien **3**, 251–278.

Mackay, D. J. (**2003**). *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, Cambridge, UK).

Miller, J. D. (**1989**). "Auditory-perceptual interpretation of the vowel," J. Acoust. Soc. Am. **85**, 2114–2133.

McGowan, R. S. (**2006**). "Perception of synthetic vowel exemplars of 4 year old children and estimation of their corresponding vocal tract shapes," J. Acoust. Soc. Am. **129**, 2850–2858.

Monsen, R. B., and Engebretson, A. M. (**1983**). "The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction," J. Speech Hear. Res. **36**, 89–97.

Patterson, R. D., van Dinther, R., and Irino, T. (**2007**). "The robustness of bio-acoustic communication and the role of normalization," Proceedings of the 19th International Congress on Acoustics, Madrid, September, p. 07–011.

Peterson, G. E., and Barney, H. I. (**1952**). "Control methods used in the study of vowels," J. Acoust. Soc. Am. **24**, 75–184.

Potter, R. K., and Steinberg, J. C. (**1950**). "Toward the specification of speech," J. Acoust. Soc. Am. **22**, 807–820.

Roweis, S., and Ghahramani, Z. (**1999**). "A unifying review of linear Gaussian models," Neural Comput. **11**, 305–345.

Umesh, S., Bharath Kumar, S. V., Vinay, M. K., Sharma, R., and Sinha, R. (**2002**). "A simple approach to non-uniform vowel normalization," *ICASSP*, Orlando, FL.

Umesh, S., Cohen, L., Marinovic, N., and Nelson, D. J. (**1999**). "Scale-transform in speech analysis," IEEE Trans. Speech Audio Process. **7**, 40–45.

Vallabha, G. K., and Tuller, B. (**2002**). "Systematic errors in the formant analysis of steady-state vowels," Speech Commun. **38**, 141–160.

Welling, M., and Ney, H. (**2004**). "Speaker adaptive modeling by vocal tract normalization," IEEE Trans. Speech Audio Process. **10**, 415–426.

Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs[a)]

Geoffrey Stewart Morrison[b)]
School of Language Studies, Australian National University, Canberra, Australian Capital Territory 0200, Australia

(Received 12 September 2008; revised 11 January 2009; accepted 15 January 2009)

Non-contemporaneous speech samples from 27 male speakers of Australian English were compared in a forensic likelihood-ratio framework. Parametric curves (polynomials and discrete cosine transforms) were fitted to the formant trajectories of the diphthongs /aɪ/, /eɪ/, /oʊ/, /aʊ/, and /ɔɪ/. The estimated coefficient values from the parametric curves were used as input to a generative multivariate-kernel-density formula for calculating likelihood ratios expressing the probability of obtaining the observed difference between two speech samples under the hypothesis that the samples were produced by the same speaker versus under the hypothesis that they were produced by different speakers. Cross-validated likelihood-ratio results from systems based on different parametric curves were calibrated and evaluated using the log-likelihood-ratio cost function (C_{llr}). The cross-validated likelihood ratios from the best-performing system for each vowel phoneme were fused using logistic regression. The resulting fused system had a very low error rate, thus meeting one of the requirements for admissibility in court.
 [DOI: 10.1121/1.3081384]

PACS number(s): 43.72.Pf [DOS] Pages: 2387–2397

I. INTRODUCTION

In determining the admissibility in court of evidence based on a scientific methodology, the Supreme Court of the United States ruled that the judge should consider multiple factors including whether the scientific methodology has been empirically tested, whether it has an acceptable error rate, and whether it has been subject to peer review and publication [Daubert versus Merrell Dow Pharmaceuticals (92-102), 509 U.S. 579 (1993)]. A number of other jurisdictions have adopted similar standards. The present paper describes a test of the use of parametric representations of vowel formant trajectories as part of a forensic-voice-comparison system. It is hoped that this paper will contribute toward meeting admissibility-of-evidence standards for this methodology. This paper is part of a larger project "Catching criminals by their voice: Combining automatic and traditional methods for optimum performance in forensic speaker identification." A phonetic-acoustic analysis of vowel formant trajectories is combined with a generative likelihood-ratio formula (this formula has previously been used in research and in court cases in conjunction with phonetic-acoustic analyses), and the results are evaluated, calibrated, and fused using procedures which have been developed for use with automatic speaker recognition.

If an acoustic variable has a between-speaker variance which is greater than its within-speaker variance then it contains information which can be used to help distinguish speakers. A number of studies have found that information relevant to speaker identity can be extracted from the dynamic spectral properties of vowels (Eriksson *et al.*, 2004; Eriksson *et al.*, 2004; Goldstein, 1976; Greisbach, 1995; Ingram *et al.*, 1996; Kasuya *et al.*, 1994; McDougall, 2004; Rodman *et al.*, 2002; Nolan and Grigoras, 2005); however, these studies were not conducted in a framework which is appropriate for the evaluation of forensic evidence. It is the position of many forensic statisticians, forensic scientists, and legal experts that the logically and legally correct method for the evaluation of forensic evidence is via likelihood ratios (LRs) (Aitken and Stoney, 1991; Aitken and Taroni, 2004; Balding, 2005; Broeders, 1999, 2004; Evett, 1998; Friedman, 1996; González-Rodríguez *et al.*, 2007; Hodgson, 2002; Jessen, 2008; Lucy, 2005; Rose, 2002, 2003, 2005, 2006b). LRs can be used to present the probability of observing the evidence given one hypothesis versus the probability of observing the same evidence given the alternative hypothesis, for example, the probability of observing the measured acoustic differences between two speech recordings under the hypothesis that the speech in the two recordings was produced by the same speaker versus under the hypothesis that it was produced by different speakers, see Eq. (1).

$$\mathrm{LR} = \frac{p(\text{observed acoustic difference}|\text{same speaker})}{p(\text{observed acoustic difference}|\text{different speakers})}. \quad (1)$$

LRs are therefore the ratio of the conditional probabilities of two competing hypotheses, and if the prior probabilities for the hypotheses were known they could be combined

[a)]Portions of this work were presented in Morrison, Rose, and Kinoshita, "Extraction of likelihood-ratio forensic evidence from the formant trajectories of diphthongs," Acoustics '08, July 2008, and in Morrison and Kinoshita, "Automatic-type calibration of traditionally derived likelihood ratios: Forensic analysis of Australian English /o/ formant trajectories," Proceedings of Interspeech 2008, September 2008.

[b)]Electronic mail: geoff.morrison@anu.edu.au

with the LR via Bayes' theorem to produce posterior probabilities, the probabilities of the hypotheses given the evidence. The forensic expert cannot, however, provide posterior probabilities. There are both logical and legal reasons for this (see the discussion in Rose, 2002, Chap. 4): Although it may be possible to make pragmatic estimates of prior probabilities, such estimates would be based on assumptions, and, under one interpretation of the Bayesian framework, the priors are the results of the trier-of-fact's (the judge or jury depending on the legal system) initial assumptions plus changes in their belief as to the relative probability of the hypotheses as a result of all the evidence already presented. Such prior probabilities are not knowable by the forensic expert and therefore they cannot calculate posterior probabilities. Determining the posterior probabilities of the guilty versus not-guilty hypotheses is the purview of the trier-of-fact, and the forensic expert should not usurp this role (Aitken and Taroni, 2004, p. 4; Evett, 1998). The expert's task is to present the weight of one piece of evidence derived via a forensic comparison of suspect and offender samples, and the judge or jury's task is to assign priors and weigh all the evidence. It is arguably preferable that the forensic expert knows as little as possible about the circumstances of the trial and the other evidence presented so that there can be no suggestion that their testimony is influenced by anything other than a scientific evaluation of the particular samples with which they were provided.

It should be stressed that a forensic-voice-comparison system with a soft LR output is quite different from a speaker-identification system which chooses one individual from a finite number of enrolled voices, and is also quite different from a speaker-verification system which determines whether a questioned voice differs from a specific enrolled voice by less than a certain threshold (Rose, 2006b).

A few studies have conducted LR analyses using information about the dynamic properties of vowel formants. Rose (2006c) and Rose *et al.* (2006) used the formant values from the initial and final targets of Australian-English diphthongs, and Kinoshita and Osanai (2006) used the initial-target values and the slope of the formants between the initial and final targets. Simple parametrizations of vowel inherent spectral change such as dual-target or target-plus-slope models, which each use two parameters, have been found to be adequate for vowel-phoneme identification (Gottfried *et al.*, 1993; Nearey and Assmann, 1986), and more sophisticated curve-fitting models do not appear to outperform simple models (Hillenbrand *et al.*, 2001). This suggests that if only simple cues such as initial and final formant values are necessary for signaling phoneme identity, then speakers may have considerable freedom in the path taken between the initial and final formant values. If the constraints on formant trajectories are relatively lax with respect to vowel-phoneme identity, then with respect to speaker identity there may be considerable information contained in the details of formant trajectories. Differences in physiology and idiosyncracies in the use of motor commands may mean that different individuals consistently produce different formant trajectories between the beginning and end of the same vowel phoneme (see Nolan, 1997).

McDougall (2006) and McDougall and Nolan (2007) conducted a series of experiments in which they fitted polynomial curves to the formant trajectories of sonorant sequences in three different data sets. Polynomials offer relatively low dimension parametrizations of formant trajectories (e.g., three parameters for quadratic and four parameters for cubic) which can capture more information about the shape of the trajectories than is available from simpler two-parameter models; however, the analyses were not conducted in a LR framework. Rose *et al.* (2006) performed a LR analysis using initial and final target values from Australian-English /aɪ/ tokens. Morrison (2008) reanalyzed the recordings from Rose *et al.* (2006) and obtained substantially better results by fitting polynomial curves to the formant trajectories and using the estimated coefficient values to calculate LR. Based on evaluation via Tippett plots, he concluded that the best results were achieved using cubic polynomials fitted to equalized-duration hertz-scaled formant trajectories. Morrison and Kinoshita (2008) conducted LR analyses using estimated coefficient values from parametric curves fitted to formant trajectories from Australian English /oʊ/ tokens. The parametric curves were second- and third-order (quadratic and cubic) polynomials and second- and third-order discrete cosine transforms. Results were evaluated using the log-LR cost (C_{llr}, Brümmer and du Preez, 2006), and again the conclusion was that the best results were achieved with cubic polynomials fitted to equalized-duration hertz-scaled formant trajectories. The current paper expands on the latter two papers by testing LR forensic-voice-comparison systems using parametric curves (polynomials and discrete cosine transforms) fitted to five vowel phonemes, /aɪ/, /eɪ/, /oʊ/, /aʊ/, and /ɔɪ/, produced by Australian-English speakers.[1] Systems based on different versions of parametric curve fitting are tested via cross-validation, calculating LR for non-contemporaneous same-speaker pairs of recordings and for different-speaker pairs of recordings. The cross-validated LR results obtained from the best-performing system for each vowel phoneme are fused to produce an overall LR for each cross-validated same-speaker and different-speaker comparison. Results are evaluated using the log-LR cost.

II. METHODOLOGY

A. Data

Data consisted of laboratory-quality audio recordings extracted from a corpus of recordings which was previously described in Kinoshita and Osanai (2006). Portions of this corpus were previously analyzed in Kinoshita and Osanai (2006), Rose *et al.* (2006), Morrison (2008), and Morrison and Kinoshita (2008).[2] The recordings were of 27 male speakers of Australian-English aged between 19 and 63 with a median of 39. The speaker was shown a card with a word written on it, for example, "high," the card was removed, and the speaker responded by saying a sentence of the form "High, H-I-G-H spells high." Table I provides a list of the words recorded and analyzed in the present study. Each speaker was recorded on two separate occasions separated by approximately 2 weeks. Within each session, written prompts were presented in random order, with each word

TABLE I. Words from the Australian English corpus which were analyzed in the present study.

	Consonant context					
Vowel	/h_/	/h_t/	/h_d/	/b_/	/b_t/	/b_d/
/aɪ/	High	Height	Hide	Buy	Bite	Bide
/eɪ/	Hay	Hate	Haydes	Bay	Bait	Spade
/oʊ/	Hoe	Hote	Hoed	Bow	Boat	Bode
/aʊ/	How	⋯	How'd	Bough	Bout	Bowed
/ɔɪ/	(Coy)	Hoytes	⋯	Boy	⋯	Buoyed

presented twice. Recordings were made using a Sony ECM-MS907 microphone and an Edirol R-1 recorder, with the signal digitized at 44.1 kHz.

B. Acoustic analysis

The first and final words from each sentence were analyzed. The beginning and end of each vowel were manually marked.[3] The trajectories of the first three formants (F1, F2, and F3) of each vowel were tracked using the formant tracking procedure outlined in Nearey *et al.* (2002): The number of linear-predictive-coding coefficients was fixed at 9. Measurements were obtained every 2 ms using a 100 ms wide power-four-cosine window. Formants were tracked using the algorithm described in Markel and Gray (1976). The formants were tracked eight times using eight different cutoff values for F3 (range 2500–4000 Hz). Each of the eight formant-track sets was visually displayed overlain on a spectrogram. Fundamental frequency was measured using the algorithm of Boersma (1993), and intensity was also measured. The measured intensity, fundamental frequency, and formant frequencies were used to synthesize a vowel. The researcher could listen to the original vowel and a synthesized vowel based on any desired track set. On the basis of visual and auditory inspection, the researcher selected what he judged to be the best formant-track set. The researcher also had the option of manually editing formant tracks and of adjusting parameters for fundamental frequency measurement.[4]

C. Parametric curve fitting

Two types of parametric curves were fitted to the formant trajectories, polynomials and discrete cosine transforms (DCTs). Earlier studies have found that DCTs are effective for vowel-phoneme discrimination and can outperform polynomials (Watson and Harrington, 1999; Zahorian and Jagharghi, 1991, 1993). A DCT is similar to a discrete Fourier transform in that it approximates a complex curve by summing a series of simple functions. DCTs are designed to be fitted to curves of finite length. The simple functions used to construct DCTs are a series of cosines with a set sequence of frequencies. The frequencies and phases are constrained so that each component has a maximum or minimum at each of the beginning and end of the complex curve. Coefficient values specify the amplitude of each cosine component (the coefficients are scaled by a factor related to the length of the signal). The zeroth DCT coefficient corresponds to the offset, the first DCT coefficient corresponds to the amplitude of a half period of a cosine, the second corresponds to the amplitude of a whole period of a cosine, the third corresponds to the amplitude of one-and-a-half periods of a cosine, etc. The formulas for constructing polynomials and DCT curves from their coefficient values are given in Eqs. (2) and (3), respectively:

$$y(x) = \sum_{k=0}^{K} \beta_k x^k, \tag{2}$$

$$y(x) = \beta_0 \frac{1}{\sqrt{N}} + \sum_{k=1}^{K} \beta_k \frac{2}{\sqrt{N}} \cos\left(\frac{(2x+1)\pi k}{2N} \right), \tag{3}$$

where $y(x)$ is the amplitude of the curve at time-point x (for DCT calculations, x values are sample frame numbers $\{0, \ldots, N-1\}$), N is the number of time points, β_k are coefficients of order k, and K is the order of the curve (total number of coefficients less 1). For DCTs the maximum value for K is $N-1$.

A priori we do not know which parametric representation of formant trajectories will produce the best results, nor do we know whether certain scales for time and frequency will produce better results than other scales; therefore, a factorial combination of curve types, orders, and scales was tested: First, second, and third degree (linear, quadratic, and cubic) polynomials and first, second, and third degree DCTs were fitted to the trajectory of each of the three formants of each vowel in the data. Curves were fitted to formant trajectories with frequency measured in hertz and in log hertz. They were also fitted to formant trajectories with time measured on the original timescale (vowels varied in duration) and on an equalized-duration timescale (all vowels linearly interpolated to have 126 points, a nominal timescale of 0–250 ms in 2 ms steps).

D. Calculation of cross-validated likelihood ratios

In general, between-speaker differences may be expected to be greater than within-speaker differences; however, it is important to note that even if two speech samples differ substantially in their acoustic properties they could have been produced by the same speaker. Speakers' voices are inherently variable and there may be substantial differences in recordings of the same speaker made only hours apart.

It is also important to note that even if two speech samples have very similar acoustic properties they could still have been produced by different speakers. If only a small

proportion of speakers in the population produce speech with acoustic properties close to some set of acoustic values A, then this set of values is atypical. If speech samples from two different speakers are drawn at random, the probability of them both having acoustic properties in the vicinity of A is low. However, if a large proportion of the population produces speech with acoustic properties close to some other set of acoustic values B, then this set of values is typical. If speech samples from two different speakers are drawn at random, the probability of them both having acoustic properties in the vicinity of B is relatively high.

Thus although in general between-speaker differences may be expected to be greater than within-speaker differences, in some instances two samples produced by different speakers could be more similar than two samples produced by the same speaker.

In order to present to the court the strength of evidence which can be extracted from speech recordings, the forensic scientist calculates a LR which quantifies the probability of obtaining the observed differences between suspect and offender samples under the hypothesis that they were produced by the same speaker versus under the hypothesis of that they were produced by different speakers (probability of evidence given hypotheses). No same-speaker versus different-speaker hard-categorization decision (probability of hypotheses given evidence) can be made.

LRs were calculated using the multivariate-kernel-density (MVKD) formula developed by Aitken and Lucy (2004).[5] The formula has previously been applied to speech data in González-Rodríguez *et al.* (2007), Ishihara and Kinoshita (2008), Morrison (2008), Morrison and Kinoshita (2008), Rose (2006c, 2007), Rose *et al.* (2006), Rose *et al.* (2008), and Zhang *et al.* (2008). The MVKD formula evaluates the similarity of two speech samples with respect to the within-speaker variation estimated from a sample of the population, and the typicality of the speech samples with respect to an estimate of the probability density of the population (based on the same sample of the population as is used to estimate the within-speaker variance). The within-speaker variance is estimated via a normal distribution, and the between-speaker population probability density is estimated via a kernel model.

Systems based on different combinations of parametric curves and time and frequency scales were tested via cross-validation. A cross-validation procedure was adopted so as to produce a more realistic picture of how each system would perform on previously unseen data, as would be the case if it were applied to a speech sample of known origin (a suspect sample) and a sample of unknown origin (an offender sample). Cross-validated LRs were calculated for each same-speaker and different-speaker pair in the data set. Each speaker's session 1 recording was compared with their own session 2 recording (a non-contemporaneous same-speaker comparison) and with every other speaker's session 2 recording (different-speaker comparisons). This produced a set of LRs where it is known whether each LR was the result of a same-speaker or a different-speaker comparison. The background sample used to estimate the within-speaker variance and the population probability density consisted of all the speakers in the data except for the speakers being compared. For example, when a same-speaker comparison was being conducted on speaker 1, then the background sample consisted of data from all speakers except speaker 1, and when a different-speaker comparison was being conducted on speakers 1 and 2, then the background sample consisted of data from all speakers except speakers 1 and 2. (Because of limitations in the MVKD formula, data from both sessions were pooled within speaker in the background sample.)

Cross-validated LRs were calculated separately for each vowel phoneme based on the estimated coefficient values from each formant trajectory obtained using each type and degree of parametric curve and using each combination of frequency and time scaling. Only combinations in which all three formant trajectories were measured on the same time and frequency scales and with the same parametric curve type and degree were considered. Because suspect and offender recordings submitted for forensic voice comparison are often recorded via telephone systems which have band-pass properties which may compromise F1, especially for vowel phonemes with intrinsically low F1, additional sets of LRs were calculated using F2 and F3 trajectories only.

E. Evaluation of system performance

The results of the cross-validated LR calculations were evaluated and calibrated using the log-LR cost, C_{llr} (Brümmer *et al.*, 2007; Brümmer and du Preez, 2006; González-Rodríguez, *et al.*, 2007; Morrison and Kinoshita, 2008; Ramos Castro, 2007; van Leeuwen and Brümmer, 2007). C_{llr} is an error metric which is independent of prior probabilities and costs, and has been adopted for use in the National Institute of Standards and Technology Speaker Recognition Evaluations (NIST SRE). The aim of calibration in forensic voice comparison is to present the information in such a way as to best aid the trier-of-fact in making appropriate decisions (Ramos Castro, 2007). Given two sets of values derived from two categories (such as same-speaker versus different-speaker comparisons) and a fixed decision boundary for classifying the values, calibration monotonically shifts and scales the values so as to produce the smallest possible error rate.

LRs are used here to express the probability of obtaining the observed acoustic difference between two speech samples given the hypothesis that the samples were produced by the same speaker over the probability of obtaining the observed acoustic difference between the two speech samples given the hypothesis that the samples were produced by different speakers. The decision boundary for LRs is 1: Values greater than 1 favor the same-speaker hypothesis and values less than 1 favor the different-speaker hypothesis (or if a logarithmic scale is used, log-LR values greater than 0 favor the same-speaker hypothesis and values less than 0 favor the different-speaker hypothesis). An ideal forensic-voice-comparison system would produce large positive log LRs for same-speaker comparisons and large negative log LRs for different-speaker comparisons. Negative log LRs from same-speaker comparisons and positive log LRs from different-speaker comparisons can be counted as errors since

they favor the contrary-to-fact hypothesis. Further, the greater the magnitude of the log LR in the wrong direction, the more egregious the error in that it is more likely to contribute toward a miscarriage of justice. If, in fact, the two speech samples being compared were produced by the same speaker, then it would be worse to present a LR of 10 000 in favor of the different-speaker hypothesis than a ratio of 10 in favor of the different-speaker hypothesis.[6] The severity of the error is captured by the C_{llr} which can be calculated using Eq. (4):

$$C_{\mathrm{llr}} = \frac{1}{2}\left(\frac{1}{N_{\mathrm{ss}}}\sum_{i=1}^{N_{\mathrm{ss}}} \log_2\left(1+\frac{1}{\mathrm{LR}_{\mathrm{ss}_i}}\right) + \frac{1}{N_{\mathrm{ds}}}\sum_{j=1}^{N_{\mathrm{ds}}} \log_2(1+\mathrm{LR}_{\mathrm{ds}_j})\right), \tag{4}$$

where N_{ss} and N_{ds} are the number of same-speaker and different speaker comparisons, and $\mathrm{LR}_{\mathrm{ss}}$ and $\mathrm{LR}_{\mathrm{ds}}$ are the LRs derived from same-speaker and different-speaker comparisons. C_{llr} is a continuous function which is small for correct LRs (same-speaker comparisons with LRs greater than 1 and different-speaker comparisons with LRs less than 1) and asymptotes toward zero as correct LRs diverge from 1, but which is large for incorrect LRs (same-speaker comparisons with LRs less than 1 and different-speaker comparisons with LRs greater than 1) and becomes exponentially large as incorrect LRs diverge from 1. C_{llr} can be decomposed into the sum of two parts: $C_{\mathrm{llr}}^{\mathrm{min}}$ is the minimum error which can be achieved for an optimally calibrated system and $C_{\mathrm{llr}}^{\mathrm{cal}}$ is the (mis)calibration error, which can be removed by optimally calibrating the system. Calibration is a monotonic shifting and scaling of the scores relative to the decision boundary and $C_{\mathrm{llr}}^{\mathrm{min}}$ is achieved using a non-parametric monotonic function (pool adjacent violators). For practical applications, systems will be calibrated using a parametric monotonic function and post-calibration C_{llr} values will therefore be greater than $C_{\mathrm{llr}}^{\mathrm{min}}$. In the present paper post-calibration C_{llr} will be used to compare the performance of different forensic-voice-comparison systems.

Pre-calibration $C_{\mathrm{llr}}(C_{\mathrm{llr}}^{\mathrm{pre}})$, post-calibration $C_{\mathrm{llr}}(C_{\mathrm{llr}}^{\mathrm{post}})$, and $C_{\mathrm{llr}}^{\mathrm{min}}$ were calculated for the sets of cross-validated LRs from each vowel phoneme. Each set of cross-validated LRs was calibrated using a linear transformation with C_{llr} as the loss function to be minimized. The C_{llr} calculations and calibration were performed using Brümmer's FOCAL toolkit.[7] Calibration was performed using cross-validation: For example, in order to calibrate the LR obtained from the same-speaker comparison of speaker 1's two recording sessions, calibration intercept and slope values were calculated using all of the LRs except those which were obtained from comparisons which included speaker 1; likewise, in order to calibrate the LR obtained from the different-speaker comparison of speakers 1 and 2, calibration intercept and slope values were calculated using all of the LRs except those which were obtained from comparisons which included speaker 1 or 2. This provides a more accurate picture of how the system might perform on previously unseen data.

F. Fusion of likelihood ratios across vowel phonemes

If LRs contain independent information, then they can be multiplied together using naïve Bayes to obtain LRs for a fused system (log LRs can be added together). In speech data and especially vowel formant data, however, substantial correlation is expected. For example, all else being equal, a speaker with a longer vocal tract will have lower average formant values for all vowel phonemes. A simple fusion procedure which takes correlation between sets of LRs into account is logistic regression (Brümmer *et al.*, 2007; González-Rodríguez *et al.*, 2007; Pigeon *et al.*, 2000; Ramos Castro, 2007). A multivariate binomial logistic regression model is fitted to a matrix of log LRs from each same-speaker and different-speaker comparison. Each row of the matrix consists of a set of log LRs from a comparison of a single pair of speech samples. The set consists of the log LRs from each of the phonemes compared, with each phoneme in a different column. The rows of the matrix are also categorically coded according to whether they were the result of a same-speaker or a different-speaker comparison. The logistic regression analysis determines the optimal hyperplane for separating the same-speaker and different-speaker comparisons. The estimated coefficient values from the logistic regression analysis are then used as weights on a linear summation of the sets of log LRs from each phoneme. Note that if equal priors are chosen for the logistic regression, then the optimization object (the deviance statistic) is the same as C_{llr}, and logistic-regression fusion therefore automatically includes calibration (Brümmer *et al.*, 2007; Brümmer and du Preez, 2006).

Logistic-regression fusion was performed using Brümmer's FOCAL toolkit.[7] Because complete separation between the training set of same-speaker and different-speaker comparisons was achieved when LRs from different phonemes were combined, the logistic regression model was trained on 999 copies of the same-speaker LRs plus 1 copy of the different-speaker LRs mislabeled as same-speaker and 999 copies of the different-speaker LRs plus 1 copy of the same-speaker LRs mislabeled as different-speaker. This prevented the algorithm from generating infinite values for coefficients (see Menard, 2002). Fusion was conducted using cross-validation, as described above for calibration, and the models were trained on the non-calibrated cross-validated LRs (results using the calibrated cross-validated LRs would be identical).

III. RESULTS AND DISCUSSION

A. Mean formant trajectories

Figure 1 provides plots of the mean equalized-duration formant tracks pooled across all tokens and speakers. Note that the pronunciation of these vowels in Australian-English may differ considerably from their pronunciation in other dialects of English such as General American and also from the pronunciations suggested by the relatively dialect neutral phonemic transcriptions used in the present paper. See Cox and Palethorpe (2007) for a brief description of Australian-English pronunciation.

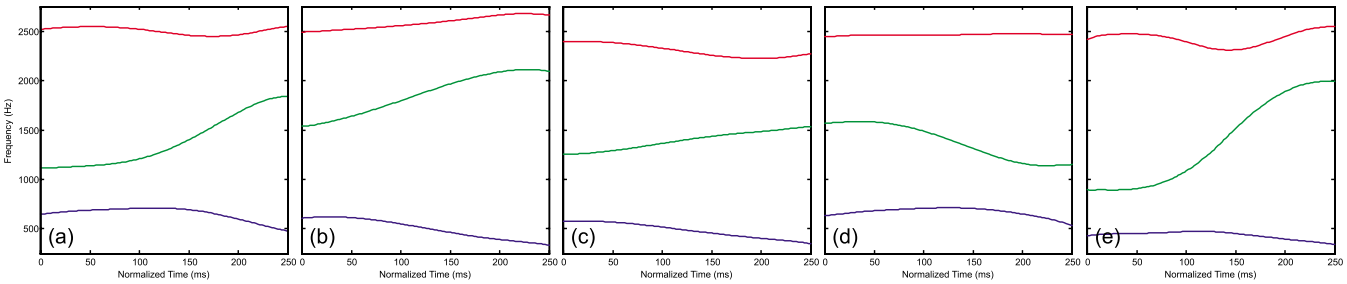


FIG. 1. (Color online) Mean equalized-duration formant tracks for Australian-English (a) /aɪ/, (b) /eɪ/, (c) /oʊ/, (d) /aʊ/, and (e) /ɔɪ/.

B. Numerical C_{llr} evaluation of forensic-voice-comparison systems

Table II provides the C_{llr} values obtained from the best-performing system for each vowel phoneme. For each phoneme the best-performing system was determined as being that based on estimated coefficient values from the type and order of parametric curve fitted to the frequency- and time-scaled formant trajectories which resulted in the lowest post-calibration C_{llr} value for the cross-validated LRs. These may be thought of as the parametric curves whose coefficients had the smallest ratio of within-speaker to between-speaker variance, and which thus extracted the greatest amount of speaker-discriminating information from the formant trajectories. Patterns of differences between different systems were often most clearly discernible for the systems based on /ɔɪ/, and to aid the discussion below Fig. 2 therefore presents a full set of pre- and post-calibration C_{llr} (C_{llr}^{pre} and C_{llr}^{post}) values covering all combinations of parametric curve type and order, and frequency and duration scales.

In general systems based on higher-order parametric curves outperformed systems based on lower-order curves. This is as would be expected if the formant trajectories are relatively complex and the highest-order curves are not modeling noise. The pattern was most clear in the C_{llr}^{post} results for /ɔɪ/ but less so for /aɪ/ or /oʊ/. There were also exceptions to this pattern: For /aʊ/ second-order polynomials generally outperformed third-order polynomials, and for /eɪ/ there was relatively little difference between the performance of first-, second-, and third-order curves for three-formant systems.

DCTs clearly outperformed polynomials for /ɔɪ/, but in general there was little difference in C_{llr}^{post} between the best DCT systems and the best polynomial systems. The best curve to fit may depend on the idiosyncracies of the shape of particular formant trajectories of particular vowel phonemes, in which case the *post hoc* procedure adopted here is preferable to *a priori* choosing a particular parametrization which may turn out to be sub-optimal.

With the exception of the two-formant systems for /aɪ/ and three-formant systems for /eɪ/, the best results were obtained using a linear-hertz frequency scale. The superiority of linear hertz over log Hz was very clear for /ɔɪ/; however, in general, there was little difference in C_{llr}^{post} between the best linear-hertz and the best log Hz systems. Again, this may depend in the idiosyncracies of the shapes of each formant trajectory and the relationships between the shapes of the different formant trajectories for each vowel phoneme.

With respect to time scaling, in many cases using equalized-duration improved system performance, but in some instances using the original non-equalized time scale produced better results. From the perspective of an articulatory model of speech production, both the time scales adopted here may appear crude. It may be that greater success can be achieved using a more sophisticated model; however, the ultimate performance of the systems presented here was very good, and the extra complexity of an articulatory-production-based model is probably not warranted at this stage.

A final observation with respect to C_{llr} values is that, in

TABLE II. Pre- and post-calibration C_{llr} values and C_{llr}^{min} values for the best-performing system for each vowel phoneme. (a) Systems based on F1, F2, and F3 trajectories. (b) Systems based on F2 and F3 trajectories.

				Scale		C_{llr} (bits)		
	Vowel	Curve type	Order	Freq.	Time	Pre-calib.	Post-calib.	C_{llr}^{min} (bits)
(a)	/aɪ/	Poly.	3	Hz	Equal.	0.953	0.113	0.061
	/eɪ/	Poly.	2	log Hz	Equal.	0.399	0.109	0.067
	/oʊ/	Poly.	3	Hz	Equal.	0.966	0.129	0.077
	/aʊ/	Poly.	2	Hz	Original	0.320	0.170	0.105
	/ɔɪ/	DCT	3	Hz	Original	0.186	0.140	0.082
(b)	/aɪ/	DCT	3	log Hz	Equal.	0.635	0.156	0.096
	/eɪ/	DCT	3	Hz	Original	0.111	0.095	0.063
	/oʊ/	Poly.	3	Hz	Equal.	0.998	0.198	0.141
	/aʊ/	DCT	3	Hz	Original	0.516	0.326	0.247
	/ɔɪ/	DCT	3	Hz	Original	0.144	0.146	0.092

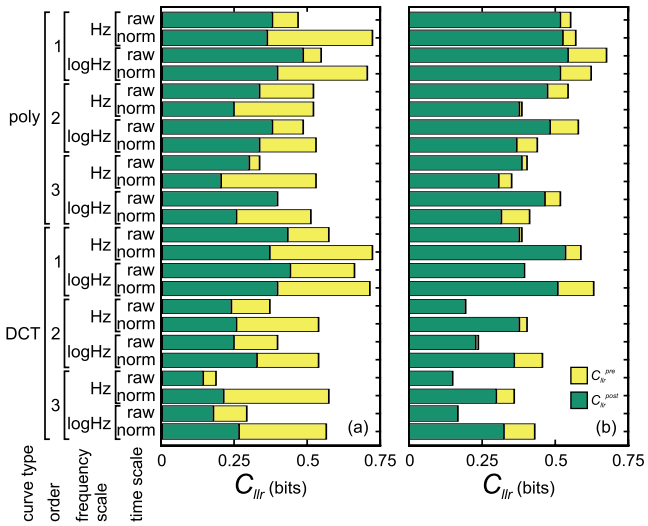


FIG. 2. (Color online) Pre- and post-calibration C_{llr} for forensic-voice-comparison systems using parametric curves fitted to the formant trajectories of Australian-English /ɔɪ/. (a) Using F1, F2, and F3 formant tracks. (b) Using F2 and F3 formant tracks.

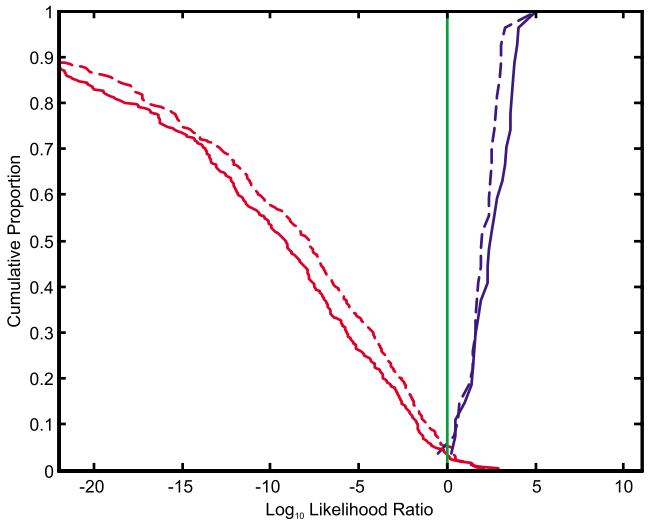


FIG. 3. (Color online) Tippett plot of the best-performing calibrated three-formant (solid line) and two-formant (dashed line) forensic-voice-comparison systems based on Australian-English /aɪ/ formant trajectories.

general, systems based on /ɔɪ/ and /eɪ/ (and especially the two-formant systems) were initially much better calibrated than systems based on other diphthongs—the difference between their pre-calibration and post-calibration C_{llr} values was smaller (for all the three-formant systems for /aɪ/, $C_{\text{llr}}^{\text{pre}}$ was at least four times greater than $C_{\text{llr}}^{\text{post}}$). This may be due to the relatively small size of the data set which was used for calculating covariance matrices. It may also be due to the fact that the Aitken and Lucy (2004) MVKD formula does not take into account all the potential sources of variance in speech data. In estimating the within-speaker variance and the population probability density, the MVKD formula does not take inter-session variance or variation due to phonetic context into account. The results based on /ɔɪ/ and /eɪ/ could therefore have been better calibrated than for the other vowels if these two vowels had less contextual variance or less inter-session variance. In the case of /ɔɪ/, this could be related to the fact that a smaller number of consonant contexts were tested for /ɔɪ/ than for other vowels (see Table I). Although the MVKD formula has been applied to speech data it was originally developed for use with trace evidence such as glass fragments. In contrast to acoustic measurements taken from speech which can vary from recording session to recording session, trace elements in panes of glass can be assumed to remain constant over time. Lacking a generative LR model which takes all the relevant sources of variance for speech data into account, calibration offers a discriminative solution for ameliorating potentially problematic raw results (see Morrison and Kinoshita, 2008).

C. Graphical Tippett-plot evaluation of forensic-voice-comparison systems

Figures 3–7 provide Tippett plots (see González-Rodríguez *et al.*, 2006; González-Rodríguez *et al.*, 2007; Rose, 2006a) of the LR results from the best-performing system for each vowel. In the Tippett plots the curves rising to the left represent the proportion of different-speaker comparisons with $\log_{10}$ LRs equal to or greater than the value indicated on the x-axis, and the curves rising to the right represent the proportion of same-speaker comparisons with $\log_{10}$ LRs equal to or less than the value indicated on the x-axis. In general, better performance is indicated by the same-speaker curve being further to the right (having more positive log LR values) and the different-speaker curve being further to the left (having more negative log LR values). The proportion and magnitude of the same-speaker curve to the left of $\log_{10}(\text{LR})=0$ and the different-speaker curve to the right of $\log_{10}(\text{LR})=0$ are indicative of errors where the calculated LRs support the contrary-to-fact hypothesis with respect to whether the comparisons were same-speaker or different-speaker comparisons. Note that the cumulative proportions in each curve are a set of discrete values determined by the number of LRs in the set as a result of the number of comparisons made. All the same-speaker curves in all the

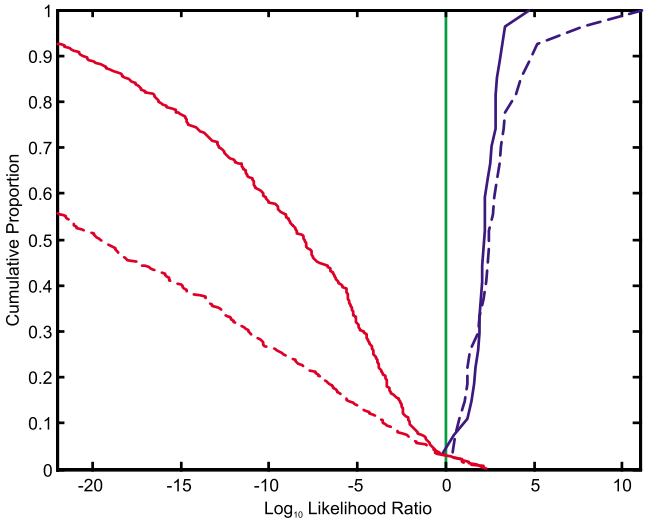


FIG. 4. (Color online) Tippett plot of the best-performing calibrated three-formant (solid line) and two-formant (dashed line) forensic-voice-comparison systems based on Australian-English /eɪ/ formant trajectories.

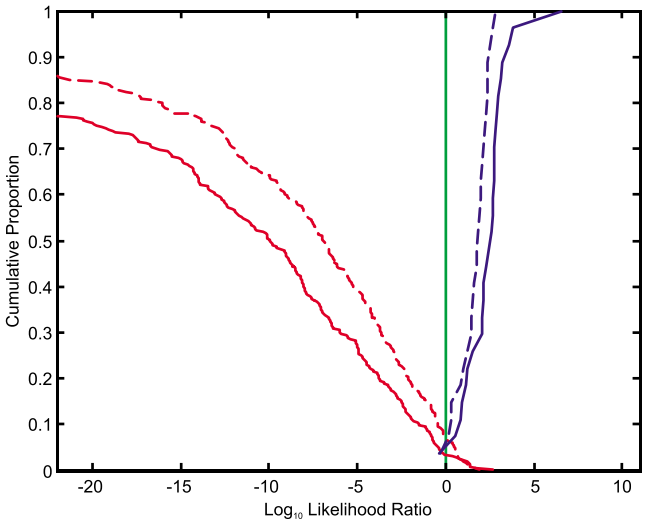


FIG. 5. (Color online) Tippett plot of the best-performing calibrated three-formant (solid line) and two-formant (dashed line) forensic-voice-comparison systems based on Australian-English /oʊ/ formant trajectories.

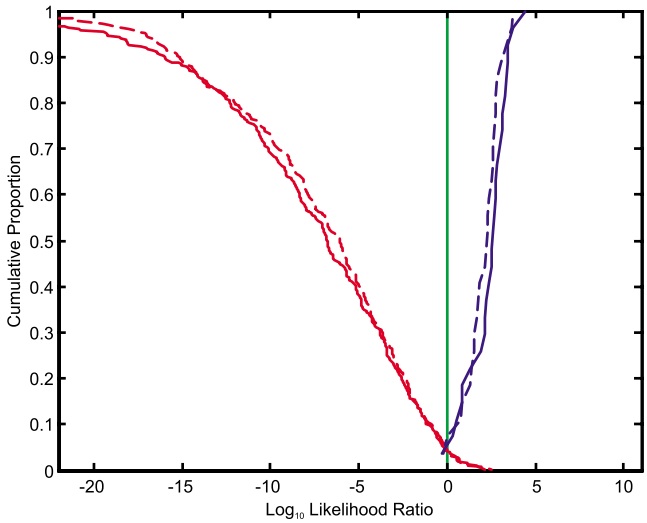


FIG. 7. (Color online) Tippett plot of the best-performing calibrated three-formant (solid line) and two-formant (dashed line) forensic-voice-comparison systems based on Australian-English /ɔɪ/ formant trajectories.

Tippett plots presented here are a result of the same number of comparisons on the same speech samples and therefore they all have the same discrete values for cumulative proportions. The same goes for the different-speaker curves. The differences between curves are therefore due to differences in the $\log_{10}$(LR) values.

For /ɔɪ/ and /aɪ/ the Tippett plots for the best-performing three-formant and two-formant systems were quite similar (Figs. 7 and 3). This suggests a high degree of correlation between the information extracted from F1 trajectories and the information extracted from F2 and F3 trajectories. For /oʊ/ and /aʊ/ the Tippett plots (Figs. 5 and 6) show a more noticeable deterioration in performance when the F1 formant trajectory is excluded (compared to the three-formant models the two-formant models resulted in Tippett plots in which the same-speaker curves have more negative log LRs and the different-speaker curves have more positive log LRs). This indicates that the F1 trajectories contained substantial extractable information about speaker identity over and above that which was extracted from the F2 and F3 trajectories. In contrast, the Tippett plot for /eɪ/ (Fig. 4) shows that substantially larger log LRs were achieved using only F2 and F3 trajectories compared to using the trajectories of all three formants. Note that $C_{\text{llr}}^{\text{post}}$ for the two-formant system was less than that for the three-formant system (Table II). This suggests that F1 trajectories for this vowel may have had relatively large within-speaker variation compared to between-speaker variation. In contrast the combination of F2 and F3 trajectories for /eɪ/ produced the best performance of any of the formant combinations tested on any vowel.

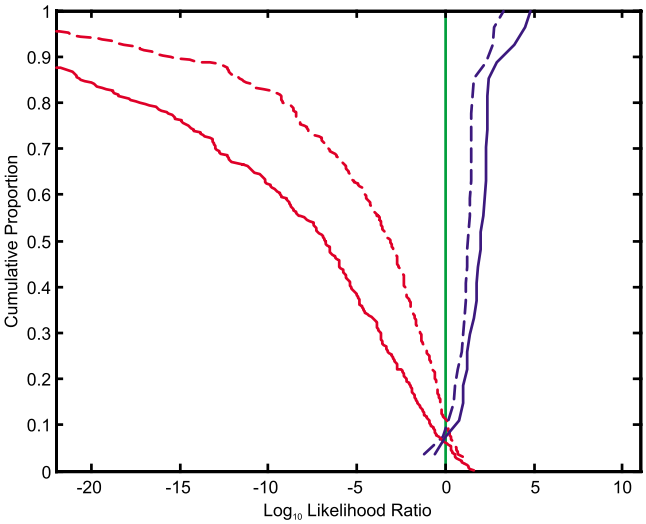


FIG. 6. (Color online) Tippett plot of the best-performing calibrated three-formant (solid line) and two-formant (dashed line) forensic-voice-comparison systems based on Australian-English /aʊ/ formant trajectories.

The cross-validated calibrated LRs from same-speaker comparisons had a low incidence of contrary-to-fact support for the different-speaker hypothesis. The worst result from a best-performing system was for the two-formant system for /aʊ/, which had a LR from a same-speaker comparison which had a magnitude of 17 in favor of the different-speaker hypothesis. For the remainder of the systems the worst LRs from same-speaker comparisons were less than 4 in favor of the different-speaker hypothesis. Given the proximity of these values to 1 they would not be interpreted as providing meaningful support for the different-speaker hypothesis. The cross-validated calibrated LRs from different-speaker comparisons had a higher incidence of contrary-to-fact support for the same-speaker hypothesis: The worst result from a best-performing system was from the three-formant system for /aɪ/, which had a LR from a different-speaker comparison which had a magnitude of 863 in favor of the same-speaker hypothesis. The best cross-validated calibrated LRs for same-speaker comparisons from the best-performing systems ranged from 726 for the two-formant system for /oʊ/ to around 147×10^9 for the two-formant system for /eɪ/.

D. Fused system

Figure 8 provides a Tippett plot of the cross-validated LRs resulting from fusing the best-performing systems for

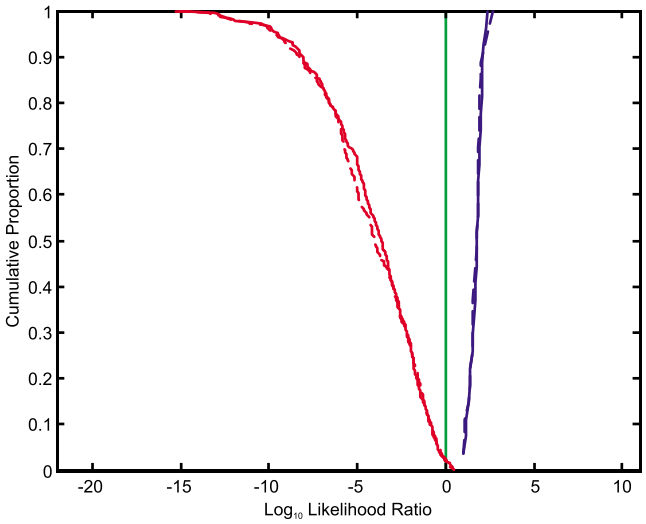


FIG. 8. (Color online) Tippett plot of the fused calibrated three-formant (solid line) and two-formant (dashed line) forensic-voice-comparison systems based formant trajectories from Australian-English /aɪ/, /eɪ/, /oʊ/, /aʊ/, and /ɔɪ/.

each vowel phoneme. The three-formant fused system included the best-performing two-formant model for /eɪ/ since that system outperformed the best-performing three-formant model for this vowel. The results for the fused two-formant and three-formant systems were very similar, suggesting that one would still obtain good results even if the bandpass properties of a telephone transmission channel rendered use of F1 trajectories impossible.

When testing via cross-validation of the training set, the fused systems achieved complete linear separation between same-speaker and different-speaker comparisons. In fact, complete linear separation could be achieved for several combinations of cross-validated LRs from only two vowel phonemes, for /aɪ/-/oʊ/, /aɪ/-/ɔɪ/, /eɪ/-/oʊ/ and for /aɪ/-/ɔɪ/.

In terms of C_{llr}, the fused systems substantially outperformed any of the single-phoneme systems: $C_{\text{llr}}^{\text{min}}$ was zero and post-fusion (post-calibration) C_{llr} was 0.057 for the three-formant system and 0.056 for the two-formant system.

There were no LRs from same-speaker comparisons which contrary-to-fact favored the different-speaker hypothesis. The smallest same-speaker LRs were 10 for the three-formant system and 11 for the two-formant system. LRs from different-speaker comparisons which contrary-to-fact supported the same-speaker hypothesis were reduced to innocuous magnitudes of less than 3. However, the largest same-speaker LRs achieved were only 229 for the three-formant system and 437 for the two-formant system.

IV. CONCLUSION

Parametric curves, polynomials and discrete cosine transforms, were fitted to the formant trajectories of tokens of /aɪ/, /eɪ/, /oʊ/, /aʊ/, and /ɔɪ/ produced by male speakers of Australian-English. The estimated coefficient values from the parametric curves were used as input to the Aitken and Lucy (2004) multivariate-LR formula, a generative formula which calculated a set of LRs for each vowel phoneme. The LRs were then calibrated and evaluated using the LR-cost function (C_{llr}). Finally the results from the best-performing system for each vowel phoneme were fused and evaluated.

There were trends indicating that DCTs generally outperformed polynomials, that third-order parametric curves generally outperformed lower-order curves, that curves fitted using a hertz frequency scale generally outperformed curves fitted using a log Hz frequency scale, and that curves fitted using an equalized-duration time scale generally outperformed curves fitted using the original non-equalized time scale. However, the parametric curve with the best performance for each vowel phoneme could only be determined on a case-by-case basis. The recommended procedure for future research and casework is therefore to proceed in the same manner as the study described here: Test a series of parametric curves using training and test data from the population sample and determine the best-performing system *post hoc* rather than *a priori* deciding a particular system which may turn out to be sub-optimal. Once the best-performing system has been chosen, that system should then be applied to obtain LRs for comparisons of suspect and offender samples (the suspect and offender samples should not be used in the process of determining the best-performing system).

Of the best-performing systems per vowel phoneme, the best performance was obtained for /eɪ/ and the poorest performance for /aʊ/ (the full rank order was /eɪ/, /aɪ/, /oʊ/, /ɔɪ/, /aʊ/). For most vowel phonemes, systems which fitted curves to the trajectories of F1, F2, and F3 outperformed systems which only fitted curves to the trajectories of F2 and F3; however, this relationship was reversed for /eɪ/. Thus although for most vowel phonemes the F1 trajectory contained extractable speaker-differentiating information in addition to that extracted from the F2 and F3 trajectories, the F1 trajectory for /eɪ/ contained confounding information. Again the implication for future research and casework is that the choice of which formant trajectories to include should be determined on a case-by-case basis.

The fused two-formant and three-formant systems produced similar results, indicating that performance would not deteriorate substantially even if F1 trajectories could not be included because of them being compromised by the bandpass properties of a telephone transmission channel. The fused systems performed extremely well in terms of having very low C_{llr} values. In cross-validated tests, complete separation was achieved between same-speaker and different-speaker comparisons. There were no same-speaker comparison results which contrary-to-fact favored the different-speaker hypothesis, and different-speaker comparison results which contrary-to-fact favored the same-speaker hypothesis were restricted to having innocuous LRs of less than 3 in favor of the same-speaker hypothesis. As far as I am aware, these are better results than have previously been achieved for any acoustic-phonetic forensic-voice-comparison system trained and tested on a database of this size.

In addition to producing very low error rates, the fusion-calibration process also resulted in what appear to be quite conservative magnitudes for LRs. The largest same-speaker LR value of 147×10^9 reported for a single phoneme is within the expected range for DNA evidence, but is suspiciously large for speech data and may be the result of poor

estimates of probability densities due to factors such as outliers or small sample sizes. For the fused system the largest same-speaker LR value was a much more conservative 437. There would therefore seem to be little danger of the expert witness overstating the degree of support for one or other of the hypotheses when using LRs generated by the fused system.

It is argued that the results presented here therefore constitute a step toward meeting the Daubert requirement that a methodology be tested and be found to have an acceptable error rate. However, additional testing will still be required before this methodology can be admitted into evidence: The present paper used data from only 27 speakers recorded under highly controlled laboratory conditions. The methodology will need to be tested using data from a larger number of speakers collected under more forensically realistic conditions, for example, extemporaneous speech from telephone conversations. In addition, since the alternative hypothesis is more likely to be "the speaker in the offender sample is someone who sounds like the speaker in the suspect sample" rather than the less-constrained "the speaker in the offender sample is someone other than the speaker in the suspect sample," the methodology should also be tested using recordings from a set of speakers who are judged to be similar by phonetically naïve listeners such as police officers and lawyers.

ACKNOWLEDGMENTS

This research was supported by Australian Research Council Grant No. DP0774115. Thanks to Yuko Kinoshita for providing the speech data. Thanks to Philip Rose, Shunichi Ishihara, Cuiling Zhang, and two anonymous reviewers for comments on an earlier draft of this paper.

[1]The author has chosen to use phonemic representations of these diphthongs which may be considered generic for English in general, and thus likely to be familiar to a larger number of readers. For consistency with the representation of the other diphthongs, the so-called phonetic diphthongs have been represented using two symbols each, /eɪ/ and /oʊ/, rather than one each, /e/ and /o/. For an alternative transcription system specifically tailored to Australian-English see Cox and Palethorpe (2007). The acoustic realizations of the five phonemes in the present study are represented graphically by the mean formant trajectories presented in Fig. 1.

[2]Of these papers, only Morrison and Kinoshita (2008) made use of a portion of the same acoustic analysis results as are reported in the present paper.

[3]A revised version of the software used to mark the vowels "SoundLabeller" is available on the author's website: http://geoff-morrison.net (last viewed January 11, 2009).

[4]A revised version of the software used to measure the vowels "FormantMeasurer" is available on the author's website: http://geoff-morrison.net (last viewed January 11, 2009).

[5]A MATLAB implementation of Aitken and Lucy's (2004) MVKD formula is available on the website of the author of the present paper: http://geoff-morrison.net (last viewed January 11, 2009).

[6]Although LR which favor the different-speaker hypothesis have values of less than 1, it is often easier to understand the presentation if the LR is inverted and a verbal clarification that it is in favor of the different-speaker hypothesis rather than the same speaker hypothesis is added, i.e., state "a likelihood ratio of 10 in favor of the different-speaker hypothesis" as opposed to stating "a likelihood ratio of 0.1."

[7]The FOCAL toolkit is available from http://niko.brummer.googlepages.com/focal (last viewed January 11, 2009). Linear calibration was achieved using the logistic-regression-fusion functions in the FOCAL toolbox with a single predictor variable.

Aitken, C. G. G., and Lucy, D. (**2004**). "Evaluation of trace evidence in the form of multivariate data," Appl. Stat. **54**, 109–122.

Aitken, C. G. G., and Stoney, D. A. (**1991**). *The Use of Statistics in Forensic Science* (Horwood, Chichester, UK).

Aitken, C. G. G., and Taroni, F. (**2004**). *Statistics and the Evaluation of Evidence for Forensic Scientists* (Wiley, Chichester, UK).

Balding, D. J. (**2005**). *Weight of Evidence for Forensic DNA Profiles* (Wiley, Chichester, UK).

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," *Proceedings of the Institute of Phonetic Sciences* (Institute of Phonetic Sciences, Amsterdam), Vol. **17**, pp. 97–110.

Broeders, A. P. A. (**1999**). "Some observations on the use of probability scales in forensic identification," Forensic Ling. **6**, 1350–1771.

Broeders, A. P. A. (**2004**). "Forensic speech and audio analysis forensic linguistics: A review 2001 to 2004," in *14th International Forensic Science Symposium Review Papers*, edited by N. Nic Daéid (Interpol, Lyon, France), pp. 170–188.

Brümmer, N., Burget, L., Cernocký, J. H., Glembek, O., Grézl, F., Karafiát, M., van Leeuwen, D. A., Matejka, P., Schwarz, P., and Strasheim, A. (**2007**). "Fusion of heterogenous speaker recognition systems in the STBU submission for the NIST SRE 2006," IEEE Trans. Audio, Speech, Lang. Process. **15**, 2072–2084.

Brümmer, N., and du Preez, J. (**2006**). "Application independent evaluation of speaker detection," Comput. Speech Lang. **20**, 230–275.

Cox, F., and Palethorpe, S. (**2007**). "Australian English," J. Int. Phonetic Assoc. **37**, 341–350.

Eriksson, E. J., Cepeda, L. F., Rodman, R. D., McAllister, D. F., Bitzer, D., and Arroway, P. (**2004**). "Cross-language speaker identification using spectral moments," in *Proceedings of FONETIK 2004: The XVIIth Swedish Phonetics Conference*, edited by P. Branderud and H. Traunmüller (Department of Linguistics, Stockholm University, Sweden), pp. 76–79.

Eriksson, E. J., Cepeda, L. F., Rodman, R. D., Sullivan, K. P. H., McAllister, D. F., Bitzer, D., and Arroway, P. (**2004**). "Robustness of spectral moments: A study using voice imitations," in *Proceedings of the Tenth Australian International Conference on Speech Sciences and Technology*, edited by S. Cassidy, F. Cox, R. Mannell, and S. Palethorpe (Australian Speech Science and Technology Association, Canberra, Australia), pp. 259–264.

Evett, I. W. (**1998**). "Towards a uniform framework for reporting opinions in forensic science casework," Sci. Justice **38**, 198–202.

Friedman, R. D. (**1996**). "Assessing evidence," Mich. Law Rev. **94**, 1810–1838.

Goldstein, U. G. (**1976**). "Speaker-identifying features based on formant tracks," J. Acoust. Soc. Am. **59**, 176–182.

González-Rodríguez, J., Drygajlo, A., Ramos, D., García-Gomar, M., and Ortega-García, J. (**2006**). "Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition," Comput. Speech Lang. **20**, 331–355

González-Rodríguez, J., Rose, P., Ramos, D., Toledano, D. T., and Ortega-García, J. (**2007**). "Emulating DNA: Rigorous quantification of evidential weight in transparent and testable forensic speaker recognition," IEEE Trans. Audio, Speech, Lang. Process. **15**, 2104–2115.

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

Greisbach, R., Esser, O., and Weinstock, C. (**1995**). "Speaker identification by formant contours," in *Studies in Forensic Phonetics*, edited by A. Braun and J.-P. Köster (Wissenschaftlicher, Trier, Germany), pp. 49–55.

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

Hodgson, D. (**2002**). "A lawyer looks at Bayes' theorem," The Aust. Law J. **76**, 109–118.

Ingram, J. C. L., Prandolini, R., and Ong, S. (**1996**). "Formant trajectories as indices of speaker identification," Forensic Ling. The Int. J. of Speech Lang. and the Law **3**, 129–145.

Ishihara, S., and Kinoshita, Y. (**2008**). "How many do we need? Exploration of the population size effect on the performance of forensic speaker classification," *Proceedings of Interspeech 2008 Incorporating SST 2008*, International Speech Communication Association, pp. 1941–1944.

Jessen, M. (**2008**). "Forensic phonetics," Lang. and Ling. Compass **2**, 671–711.

Kasuya, H., Tan, X., and Yang, C.-S. (**1994**). "Voice source and vocal tract

characteristics associated with speaker individuality," Proceedings of the Third International Conference on Spoken-Language Processing, Yokohama, pp. 1459–1462.
Kinoshita, Y., and Osanai, T. (**2006**). "Within speaker variation in diphthongal dynamics: What can we compare?" in *Proceedings of the 11th Australasian International Conference on Speech Science and Technology, Auckland, New Zealand*, edited by P. Warren and C. I. Watson (Australasian Speech Science and Technology Association, Canberra, Australia), pp. 112–117.
Lucy, D. (**2005**). *Introduction to Statistics for Forensic Scientists* (Wiley, Chichester, UK).
Markel, J. D., and Gray, A. H. (**1976**). *Linear Prediction of Speech* (Springer-Verlag, Berlin).
McDougall, K. (**2004**). "Speaker-specific formant dynamics: An experiment on Australian English /aɪ/," Int. J. of Speech, Lang. and the Law **11**, 103–130.
McDougall, K. (**2006**). "Dynamic features of speech and the characterization of speakers," Int. J. of Speech, Lang. and the Law **13**, 89–126.
McDougall, K., and Nolan, F. (**2007**). "Discrimination of speakers using the formant dynamics of /u/ in British English," in *Proceedings of the 16th International Congress on Phonetic Sciences, Saarbrücken*, edited by J. Trouvain and W. J. Barry (Universität des Saarlandes, Saarbrücken, Germany), pp. 1825–1828.
Menard, S. (**2002**). *Applied Logistic Regression Analysis*, 2nd ed. (Sage, Thousand Oaks, CA).
Morrison, G. S. (**2008**). "Forensic voice comparison using likelihood ratios based on polynomial curves fitted to the formant trajectories of Australian English /aɪ/," Int. J. of Speech, Lang. and the Law **15**, pp. 247–264.
Morrison, G. S., and Kinoshita, Y. (**2008**). "Automatic-type calibration of traditionally derived likelihood ratios: Forensic analysis of Australian English /o/ formant trajectories," *Proceedings of Interspeech 2008 Incorporating SST 2008*, International Speech Communication Association, pp. 1501–1504.
Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of vowel inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.
Nearey, T. M., Assmann, P. F., and Hillenbrand, J. M. (**2002**). "Evaluation of a strategy for automatic formant tracking," J. Acoust. Soc. Am. **112**, 2323.
Nolan, F. (**1997**). "Speaker recognition and forensic phonetics," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell, Oxford), pp. 744–767.
Nolan, F., and Grigoras, C. (**2005**). "A case for formant analysis in forensic speaker identification," Int. J. of Speech, Lang. and the Law **12**, 144–173.
Pigeon, S., Druyts, P., and Verlinde, P. (**2000**). "Applying logistic regression to the fusion of the NIST'99 1-speaker submissions," Digit. Signal Process. **10**, 237–248.
Ramos Castro, D. (**2007**). "Forensic evaluation of the evidence using automatic speaker recognition systems," Ph.D. dissertation, Universidad Autónoma de Madrid, Madrid, Spain.
Rodman, R., McAllister, D., Bitzer, D., Cepeda, L., and Abbitt, P. (**2002**). "Forensic speaker identification based on spectral moments," Int. J. of Speech, Lang. and the Law **9**, 22–43.
Rose, P. (**2002**). *Forensic Speaker Identification* (Taylor & Francis, London).
Rose, P. (**2003**). *Expert Evidence, Issue 99: The Technical Comparison of Forensic Voice Samples* (Thomson Lawbook, Sydney, Australia).
Rose, P. (**2005**). "Forensic speaker recognition at the beginning of the twenty-first century: An overview and a demonstration," Aus. J. of Forensic Sci. **37**, 49–72.
Rose, P. (**2006a**). "Accounting for correlation in linguistic-acoustic likelihood ratio-based forensic speaker discrimination," in Speaker and Language Recognition Workshop, 2006. IEEE Odyssey 2006, edited by K. Berkling and P. A. Torres-Carrasquillo, pp. 1–8.
Rose, P. (**2006b**). "Technical forensic speaker recognition: Evaluation, types and testing of evidence," Comput. Speech Lang. **20**, 159–191.
Rose, P. (**2006c**). "The intrinsic forensic discriminatory power of diphthongs," in *Proceedings of the 11th Australasian International Conference on Speech Science and Technology, Auckland, New Zealand*, edited by P. Warren and C. I. Watson (Australasian Speech Science and Technology Association, Canberra, Australia), pp. 64–69.
Rose, P. (**2007**). "Forensic speaker discrimination with Australian English vowel acoustics," in *Proceedings of the 16th International Congress on Phonetic Sciences, Saarbrücken*, edited by J. Trouvain and W. J. Barry (Universität des Saarlandes, Saarbrücken, Germany), pp. 1817–1820.
Rose, P., Kinoshita, Y., and Alderman, T. (**2006**). "Realistic extrinsic forensic speaker discrimination with the diphthong /aɪ/," in *Proceedings of the 11th Australasian International Conference on Speech Science and Technology, Auckland, New Zealand*, edited by P. Warren and C. I. Watson (Australasian Speech Science and Technology Association, Canberra, Australia), pp. 329–334.
Rose, P., Kinoshita, Y., and Ishihara, S. (**2008**). "Beyond the long-term mean: Exploring the potential of F0 distribution parameters in traditional forensic speaker recognition," Proceedings of the Odyssey Speaker and Language Recognition Workshop 2008, Stellenbosch, South Africa, pp. 329–334.
van Leeuwen, D. A., and Brümmer, N. (**2007**). "An introduction to application-independent evaluation of speaker recognition systems," in *Speaker Classification I: Selected Projects*, edited by C. Müller (Springer, Heidelberg, Germany), pp. 330–353.
Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence of dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am. **106**, 458–468.
Zahorian, S. A., and Jagharghi, A. J. (**1991**). "Speaker normalization of static and dynamic vowel spectral features," J. Acoust. Soc. Am. **90**, 67–75.
Zahorian, S. A., and Jagharghi, A. (**1993**). "Spectral-shape features versus formants as acoustic correlates for vowels," J. Acoust. Soc. Am. **94**, 1966–1982.
Zhang, C., Morrison, G. S., and Rose, P. (**2008**). "Forensic speaker recognition in Chinese: A multivariate likelihood ratio discrimination on /ɪ/ and /y/," *Proceedings of Interspeech 2008 Incorporating SST 2008*, International Speech Communication Association, pp. 1937–1940.

Absolute pitch among students in an American music conservatory: Association with tone language fluency

Diana Deutsch[a] and Kevin Dooley
Department of Psychology, University of California, San Diego, La Jolla, California 92093

Trevor Henthorn
Department of Music, University of California, San Diego, La Jolla, California 92093

Brian Head
Thornton School of Music, University of Southern California, Los Angeles, California 90089

(Received 6 October 2008; revised 15 January 2009; accepted 18 January 2009)

Absolute pitch (AP), the ability to name a musical note in the absence of a reference note, is extremely rare in the U.S. and Europe, and its genesis is unclear. The prevalence of AP was examined among students in an American music conservatory as a function of age of onset of musical training, ethnicity, and fluency in speaking a tone language. Taking those of East Asian ethnicity, the performance level on a test of AP was significantly higher among those who spoke a tone language very fluently compared with those who spoke a tone language fairly fluently and also compared with those who were not fluent in speaking a tone language. The performance level of this last group did not differ significantly from that of Caucasian students who spoke only nontone language. Early onset of musical training was associated with enhanced performance, but this did not interact with the effect of language. Further analyses showed that the results could not be explained by country of early music education. The findings support the hypothesis that the acquisition of AP by tone language speakers involves the same process as occurs in the acquisition of a second tone language. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3081389]

PACS number(s): 43.75.Cd [NHF] Pages: 2398–2403

I. INTRODUCTION

Absolute pitch (AP)—the ability to name a musical note in the absence of a reference note—is very rare in North America and Europe, with an estimated prevalence of less than 1 in 10,000 (Bachem, 1955; Profita and Bidder, 1988). This ability is not necessarily accompanied by superior performance on other auditory or musical tasks (Bachem, 1937; Rakowski, 1972; Lockhead and Byrd, 1981; Rakowski and Morawska-Büngeler, 1987; Miyazaki, 1992, 2004; Burns and Campbell, 1994; Miyazaki and Rakowski, 2002; Fujisaki and Kashino, 2002), and its genesis is unclear.

Perhaps the most surprising feature of AP is its rarity. Most people readily identify hundreds of musical passages by name, yet AP involves choosing between only the 12 notes of the chromatic scale: C, C#, D, and so on. Furthermore, this ability is rare even among professional musicians, who spend tens of thousands of hours reading musical scores and playing the notes they read. Indeed, the vast majority of documented attempts to train adult musicians to acquire AP have been strikingly unsuccessful (Takeuchi and Hulse, 1993; Ward, 1999; Rakowski and Miyazaki, 2007).

Reports from possessors of AP indicate that the earlier the onset of musical training the higher the probability of acquiring this ability (Bachem, 1955; Miyazaki, 1988; Profita and Bidder, 1988). Large-scale studies involving both surveys (Baharloo *et al.*, 1998) and direct tests (Deutsch *et al.*, 2006) have found AP to be most prevalent among those who had begun musical training by ages 4–5, less prevalent between ages 6–8, and very rare after age 9. This association between the possession of AP and age of onset of musical training has led to the surmise that a critical period beginning at around age 2 or 3 is involved in its acquisition (see, for example, Takeuchi and Hulse, 1993). However since it is impractical to initiate formal musical training at an earlier age, the possibility arises that, in principle, this critical period might extend down to infancy. Indeed, Saffran and Griepentrog (2001) found that infants were able to perform a perceptual learning task that required referring to the absolute pitches of tones.

There is an intriguing parallel between the time frames involved in acquiring speech and language, on the one hand (Kuhl *et al.*, 1992), and AP on the other. Concerning first language acquisition, there are rare documented cases of children who had been socially isolated early in life and were later placed in a normal environment; these children were found unable to acquire normal language (Lane, 1976; Curtiss, 1977). Studies of recovery of speech following brain injury have indicated the same time frame: The prognosis for recovery has been found most positive for those whose injury occurred before age 6, less positive between ages 6 and 8, and very poor after puberty (Dennis and Whitaker, 1976; Bates, 1992; Duchowny *et al.*, 1996; Doupe and Kuhl, 1999). Second language acquisition beyond early childhood occurs far more readily than does acquisition of a first language, presumably because the basic underlying circuitry has already been established. However, acquisition of a second

[a]Electronic mail: ddeutsch@ucsd.edu

 0001-4966/2009/125(4)/2398/6/$25.00

language is also subject to a critical period: It occurs extremely readily in infancy (Kuhl *et al.*, 2003), with proficiency in its acquisition declining with increasing age of initial exposure to the second language, reaching a plateau in adulthood (Lennenberg, 1967; Johnson and Newport, 1989; Newport, 1990).

Given the similarity between the time frames for acquiring speech, on one hand, and AP, on the other, it has been suggested that AP was originally packaged in with other features of speech, so that it might be expected to be heavily influenced by the speech-related critical period (Deutsch, 2002; Deutsch *et al.*, 2004, 2006). This hypothesis was first proposed based on the finding that speakers of two tone languages (Vietnamese and Mandarin) displayed a remarkably precise and stable form of AP in reciting lists of words on different days (Deutsch *et al.*, 2004). In tone languages, words take on arbitrarily different lexical meanings depending on the "tones" in which they are pronounced. Tones are defined by both their pitch heights and contours. In Mandarin, for example, the first tone is characterized as high and level, the second as mid-high and rising, the third as initially low and descending and then rising, and the fourth as high and falling. For example, the word "ma" means "mother" when spoken in the first tone, "hemp" in the second, "horse" in the third, and a reproach in the fourth. When, therefore, speakers of Mandarin hear the word "ma" spoken in the first tone, and attribute the meaning "mother," or when they hear "ma" spoken in the second tone and attribute the meaning "hemp," they are using pitch (along with other features of speech) in the process of attributing a verbal label to the sound. Analogously, when people with AP hear the tone F# and identify it as "F#," or hear the note E and identify it as "E," they are also using pitch to attribute a verbal label to the sound.

On this line of reasoning, it was hypothesized that in cultures where tone languages are spoken, infants generally acquire AP for the tones of their language during the critical period in which they acquire other features of their native language (Deutsch, 2002). When they reach the age at which they can begin musical training, they acquire AP for musical tones in the same way as they would acquire the tones of a second tone language (see, for example, Wayland and Guion, 2004). For such individuals, therefore, the acquisition of AP for musical tones should also be subject to a critical period, but its overall prevalence should be considerably higher than for those who had instead acquired a nontone language in infancy and so would not have had the opportunity to develop the basic underlying circuitry at that time.

In a study designed to evaluate this hypothesis, two large groups of music students were administered a direct on-site test of AP, without self-selection from among the target populations. The first group comprised students at the Central Conservatory of Music in Beijing; these were all tone language speakers. The second group comprised students at Eastman School of Music who were nontone language speakers. The data from both groups displayed large and orderly effects of age of onset of musical training, and importantly, the speakers of tone language far outperformed those of nontone language for all levels of age of onset of training (Deutsch *et al.*, 2006).

The above findings are in accordance with the hypothesis that the acquisition of AP is subject to a speech-related critical period. However, the findings are also consistent with a genetic explanation for the difference between the two groups in prevalence of AP. The present study was designed to evaluate the hypothesis that AP is influenced by a speech-based critical period, holding genetic factors constant.

We compared four groups of subjects. The first group, termed *nontone*, consisted of those who were Caucasian and spoke only nontone language. The remaining subjects were of East Asian ethnic heritage, with both parents speaking an East Asian tone language. These subjects were assigned to three groups in accordance with their responses to a questionnaire. Those in group *tone very fluent* responded that they spoke an East Asian tone language very fluently; those in group *tone fairly fluent* responded that they spoke an East Asian tone language fairly fluently; those in group *tone nonfluent* responded that they were not fluent in speaking a tone language. By comparing the prevalence of AP in these four groups, we evaluated the relative contributions of ethnic heritage and language to the probability of acquiring AP.

II. METHOD

A. Procedure

Subjects were administered a test for AP that was identical to that given in the study of Deutsch *et al.* (2006). They were presented successively with the 36 tones spanning three octaves from C_3 (131 Hz) to B_5 (988 Hz) and were asked to write down the name of each tone (C, C#, D, and so on) when they heard it. In order to minimize the use of relative pitch as a cue, all successively presented tones were separated by an interval larger than an octave. The stimuli were piano tones that were generated on a Kurzweil K2000 synthesizer tuned to A_4=440 Hz. The tones were all 500 ms in duration and were presented in three blocks of 12, with 4.25-s intervals between onsets of tones within a block and 39-s pauses between blocks. The three test blocks were preceded by a practice block in which four successive tones were presented. The subjects were given no feedback, either during the test blocks or during the practice blocks. The tones were presented via a compact disk player, amplifier, and two loudspeakers.

Following the test, the subjects filled out a questionnaire that enquired into their music education, ethnic heritage, where they and their parents had lived, languages they and their parents spoke, and how fluently they spoke each language. Specifically concerning linguistic fluency, the subjects chose between the following options: (a) I speak the language very fluently, (b) I speak the language fairly fluently, and (c) I can understand the language, but don't speak it fluently.

B. Subjects

There were 203 subjects in this study. These were 110 males and 93 females, with average age 19.5 years (range

18–33 years). 162 subjects were first- or second-year students taking a required course at Thornton School of Music at the University of Southern California (USC); the remaining 41 subjects were members of the USC Community Orchestra. All those who were invited to take the test agreed to do so, so there was no self-selection of subjects in this experiment. All subjects were free of hearing problems, as indicated by self-report.

Based on their responses to the questionnaire, the subjects were divided into four groups. Those in group *nontone* were Caucasian and spoke only nontone language fluently. These were $n=176$, 98 males and 78 females, with average age 19.4 years (range 18–33 years). The remaining subjects were all of East Asian (Chinese or Vietnamese) ethnic heritage, with both parents primarily speaking an East Asian tone language. These were assigned to three further groups in accordance with their responses to the questionnaire. Those in group *tone very fluent* reported that they spoke an East Asian tone language "very fluently." These were $n=15$, 6 males and 9 females, with average age 20.3 years (range 18–28 years). Those in group tone *fairly fluent* reported that they spoke a tone language "fairly fluently." These were $n=7$, 4 males and 3 females, with average age 21.3 years (range 18–29 years). Those in group *tone nonfluent* responded "I can understand the language, but don't speak it fluently." These were $n=5$, 2 males and 3 females, with average age 18.6 years (range 18–19 years).

Each of the four groups was divided into subgroups by age of onset of musical training: One subgroup consisted of those who had begun musical training at ages 2–5. For *nontone*, $n=49$; for *tone very fluent*, $n=12$; for *tone fairly fluent*, $n=3$; and for *tone nonfluent*, $n=4$. Another subgroup consisted of those who had begun musical training at ages 6–9. For *nontone*, $n=87$; for *tone very fluent*, $n=3$; for *tone fairly fluent*, $n=4$; and for *tone nonfluent*, $n=1$. A third subgroup of nontone language speakers ($n=40$) comprised those who had begun musical training at age ≥ 10. However, all the subjects of East Asian ethnic heritage had begun musical training at age ≤ 9. For this reason, the data from the nontone language speakers with onset of musical training at age ≥ 10 are displayed but not compared statistically with that of any of the other groups.

III. RESULTS

Figure 1 shows, for each subgroup, the mean percentage correct on the test for AP. As is evident, all groups showed clear effects of age of onset of musical training: Performance levels on this test were higher for those who had begun musical training at ages 2–5 than at ages 6–9 (and also than at age ≥ 10 in the case of the nontone language speakers). This association with age of onset of musical training is in accordance with earlier findings (Baharloo *et al.*, 1998; Deutsch *et al.*, 2006).

It can also be seen that those subjects who stated that they spoke a tone language very fluently (*tone very fluent*) displayed remarkably high performance levels on this test. Indeed, their performance was far higher than that of the nontone language speakers (*nontone*). It was also far higher than that of the subjects of same ethnicity but who did not speak a tone language fluently (*tone nonfluent*). In addition, it was higher than that of the *tone fairly fluent* speakers, which was in turn higher than that of the *tone nonfluent* speakers and of the nontone language speakers (*nontone*).

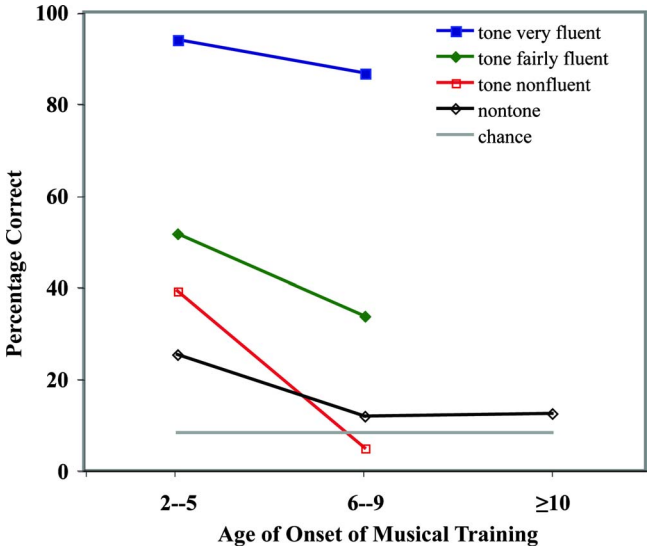


FIG. 1. (Color online) Percentage correct responses on the test of AP, as a function of age of onset of musical training and fluency in speaking a tone language. Those in groups *tone very fluent*, *tone fairly fluent*, and *tone nonfluent* were all of East Asian ethnic heritage. Those in group *nontone* were Caucasian and spoke only nontone language. The line labeled *chance* represents chance performance on the task.

To make statistical comparison between the groups, an overall test for the effect of gender was first performed and found to be nonsignificant ($F<1$); the data from males and females were therefore combined in all further analyses. A 4×2 analysis of variance (ANOVA) was performed, with group (*nontone*, *tone very fluent*, *tone fairly fluent*, and *tone nonfluent*) and age of onset of musical training (*2–5* and *6–9*) as factors. A significant effect of age of onset was found $[F(1,\ 148)=5.15, p=0.025]$. In addition, a highly significant effect of group was found $[F(3,\ 148)=35.43, p<0.001]$. The interaction between group and age of onset was nonsignificant ($F<1$).

On post hoc comparisons, the performance level was significantly higher for *tone very fluent* compared with *nontone* ($p<0.001$, one-tailed), for *tone very fluent* compared with *tone nonfluent* ($p<0.001$, one-tailed), and for *tone very fluent* compared with *tone fairly fluent* ($p<0.001$, one-tailed). The performance level for *tone fairly fluent* was significantly higher than for *nontone* ($p=0.003$) and also higher than for *tone nonfluent* (though the latter comparison was not statistically significant; $p>0.05$). The performance level for the *tone nonfluent* group did not differ significantly from that for the *nontone* group ($p>0.05$). These findings indicate strongly that the differences in performance levels between the groups were determined by language rather than ethnicity. Indeed, taking those subjects of East Asian ethnic heritage, a regression analysis found that fluency in speaking a tone language was a highly significant predictor of performance ($R^2_{\text{adj}}=0.54$, $F(2,24)=16.09$, and $p<0.001$).

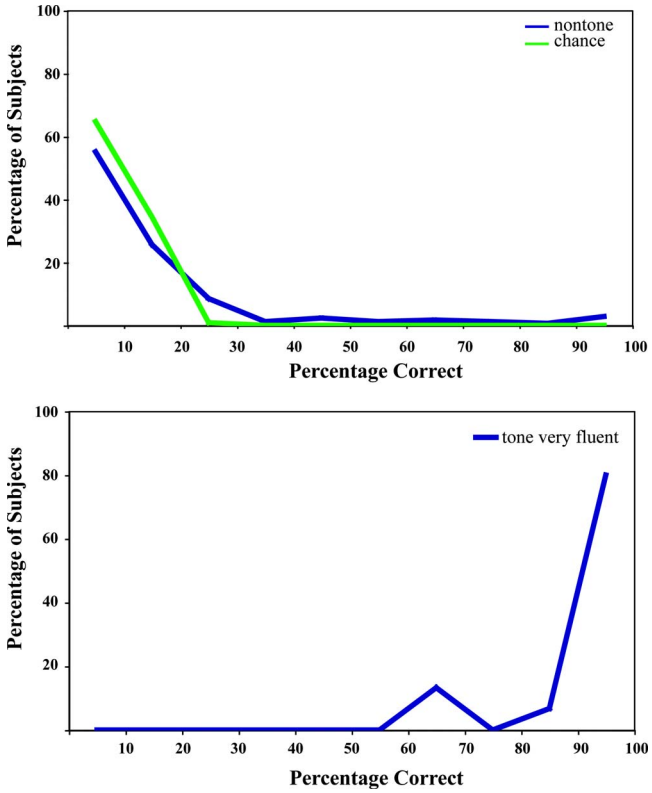


FIG. 2. (Color online) Upper graph: Relative distribution of scores of the nontone language speakers (group *nontone*), together with the hypothetical distribution of scores expected from chance performance. Lower graph: relative distribution of scores in the group *tone very fluent*.

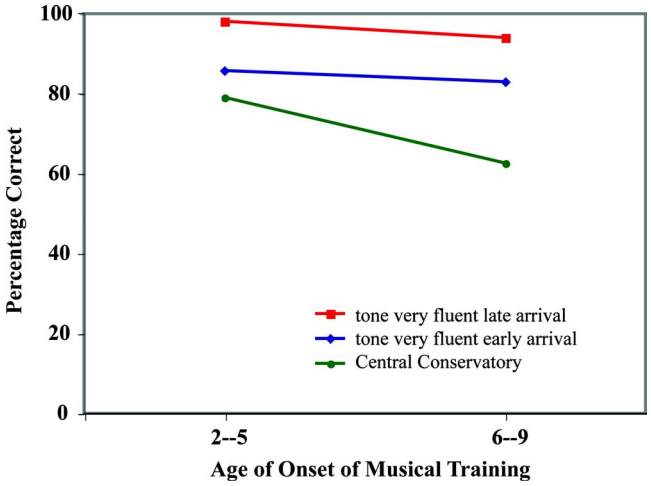


FIG. 3. (Color online) Percentage correct responses on the test of absolute pitch, as a function of age of onset of musical training, among those who were very fluent in speaking a tone language. Those in group *tone very fluent early arrival* had been born in the U.S. or had arrived in the U.S. before age 9. Those in group *tone very fluent late arrival* had arrived in the U.S. after age 9. The data from group *Central Conservatory* were obtained in an earlier study (Deutsch *et al.*, 2006), taken from subjects who had grown up in China and were tested at the Central Conservatory of Music in Beijing, China.

The strong relationship between the prevalence of AP and fluency in speaking a tone language is also reflected in the data shown in Fig. 2. The upper graph shows the relative distribution of scores of all nontone language speakers in the study, together with the hypothetical distribution of scores expected from chance performance (i.e., assuming a 1/12 chance of guessing correctly across 36 trials). We can note that there is the striking similarity between the two plots, with a very slight increase in the prevalence of AP in the 90%–100% region. The lower graph shows, in contrast, the relative distribution of scores in the *tone very fluent* group. As can be seen, the performance level of most of these subjects was in the 90%–100% region.

One issue that arises from these findings concerns the country in which the subjects received their music education. To evaluate this factor, comparison was made, taking only the *tone very fluent (TVF)* group, between those who had been born in the U.S. or who had arrived in the U.S. before age 9 (*TVF early arrivals*) and those who had arrived in the U.S. after age 9 (*TVF late arrivals*). For *TVF early arrivals*, $n=6$, with average age=19.3 (range=18–21). For *TVF late arrivals*, $n=9$, with average age=20.9 (range 18–28). A 2×2 ANOVA was performed, with group (*TVF early arrivals* vs *TVF late arrivals*) and age of onset of musical training (*2–5* vs *6–9*) as factors. The effect of age of onset of musical training was nonsignificant ($F<1$); however, this can be accounted for by a ceiling effect. The difference between the *TVF early* and *late arrivals* was also nonsignificant $[F(1,11)=2.107, p>0.05]$.

Given the remarkably high performance of the two *tone very fluent* groups, we compared their performance levels with those obtained by Deutsch *et al.* (2006) from the students at the Central Conservatory of Music in Beijing who had been administered the identical test for AP. The subjects tested in Beijing had all been born and grown up in China. In this study, we had compared the numbers of subjects in each group who obtained at least 85% correct on the test. In the present experiment, as with the other analyses we performed here, we instead employed the more sensitive measure of percentage correct for each subject. We divided the subjects from the Central Conservatory of Music into the same two categories based on age of onset of musical training as for the subjects in the present experiment, i.e., ages 2–5 and ages 6–9. The *Central Conservatory* subgroup with age of onset of musical training 2–5 consisted of 46 subjects, and the subgroup with age of onset of musical training 6–9 consisted of 34 subjects.

Figure 3 displays the average performance levels of each of these six subgroups. A 2×3 ANOVA was carried out with age of onset (*2–5* and *6–9*) and group (*TVF early arrival*, *TVF late arrival*, and *Central Conservatory*) as factors. The effect of age of onset was nonsignificant ($F<1$). However, this was presumably attributable to the ceiling effect within the TVF groups described above. Taking the *Central Conservatory* group alone, where performance was not at ceiling, the effect of age of onset was clearly significant $[F(1,78)=5.69, p=0.02]$.

Interestingly, the overall effect of group was also nonsignificant $[F(2,89)=1.76, p>0.05]$. In *post hoc* comparisons, the performance of the *TVF early arrivals* was not significantly different from either the *TVF late arrivals* ($p>0.05$) or the *Central Conservatory* group ($p>0.05$). However, the performance level of the *TVF late arrivals* was significantly higher than that of the *Central Conservatory*

group (p=0.013), and the performance level of the *TVF early arrivals* was also higher as a trend than that of the *Central Conservatory* group. This pattern of results shows that the higher performance level among the *tone very fluent* USC students in this study cannot be attributed to country of early music education.

IV. DISCUSSION

The findings presented here indicate strongly that the enhanced prevalence of AP among tone language speakers found by Deutsch *et al.* (2006) was language-related rather than genetic in origin. The present findings further show that when degree of fluency in speaking a tone language is held constant, differences in prevalence of AP among the various groups cannot be explained by the country in which the subjects had received their music education.

Specifically, with respect to the genetic issue, we found, among subjects of the same ethnic heritage, significant differences in performance level depending on degree of fluency in speaking a tone language. In addition, we found no significant difference in performance level between the (genetically East Asian) *tone nonfluent* group and the (genetically Caucasian) *nontone* group. With respect to the surmise that the differences we obtained might be related to country of early music education, we found no significant difference in overall performance level between the *tone very fluent (TVF)* subjects and those tested in the study of Deutsch *et al.* (2006) who were students at the Central Conservatory of Music in Beijing and who had all received their music education in China. Furthermore, we found no significant difference between the (USC) *TVF early arrivals* and *TVF late arrivals*, who all spoke a tone language very fluently. Indeed, the prevalence of AP was higher as a trend among the *TVF early arrivals*, who had received their music education in the U.S., than among the *Central Conservatory* students, who had received their music education in China. This difference might be accounted for by the fact that many of the *Central Conservatory* students would have had extensive experience with Asian musical scales, and this could have interfered to some extent with the note naming task employed here. It is further interesting to observe that there was a trend for the *TVF late arrivals* to perform better than the *TVF early arrivals*, and we may conjecture that this superior performance resulted from their more extensive experience with tone language. Furthermore, since the *TVF late arrivals* were studying at USC we may conjecture that they would have been more immersed in the tradition of Western tonal music, and so with Western musical scales, than were the *Central Conservatory* students; this could account for the higher performance levels of the *TVF late arrivals* compared with the *Central Conservatory* group.

It is also interesting to note that, taking the subjects of Chinese or Vietnamese ethnic heritage, those who spoke a tone language fluently were considerably more numerous than were those who were not fluent in speaking a tone language, and further that for all subjects of Chinese or Vietnamese ethnic heritage (including those born in the U.S.) both parents spoke a tone language fluently. This contrasts with the assumption that is frequently made that East Asian students in music schools in the U.S. would include many who speak only English. The present findings show that using the criterion of "ethnicity" without considering linguistic factors could lead to incorrect conclusions about the genesis of AP [see also the survey study by Gregersen *et al.* (2001) and the Comment on this study by Henthorn and Deutsch (2007)].

It is further interesting to note that all subjects in the *tone very fluent*, *tone fairly fluent*, and *tone nonfluent* groups would have been exposed to tone language in infancy. This shows that exposure to tone language during the critical period for language acquisition is not, in itself, sufficient to produce an advantage for the acquisition of AP. Our findings are, however, in accordance with those of others showing substantial first language attrition in children who, having initially been exposed to one language, later primarily speak a different language (see, for example, Kaufman and Aronoff, 1991; Nicoladis and Grabois, 2002; Ventureyra *et al.*, 2004). We must conclude from our present findings that continuity in speaking a tone language is an important factor in the probability of acquiring this ability.

The present findings leave open the question of the genesis of AP among individuals who speak only nontone language. On the present line of reasoning, there would be an advantage to those who are born into families of musicians since they would have had the opportunity to associate musical tones with their names very early in life—frequently before they begin formal musical training. This could account for findings of familial aggregation of AP (Bachem, 1955; Profita and Bidder, 1988; Baharloo *et al.*, 1998) though such findings are also in accordance with genetic explanations. It has further been conjectured (Deutsch, 2002, 2006; Deutsch *et al.*, 2004) that those who are not born into a strongly musical environment and nevertheless acquire AP may have a language-related critical period of unusually long duration, so that it extends to the age at which they can begin taking music lessons. Such an extended critical period might well have a genetic basis. The finding of an enhanced leftward asymmetry of the planum temporale among Western AP possessors (Schlaug *et al.*, 1995; Zatorre *et al.*, 1998; Keenan *et al.*, 2001) would indicate that hardwired, and so possibly genetic, factors could be involved here. Nevertheless, the factors determining the acquisition of AP by nontone language speakers remain unresolved.

ACKNOWLEDGMENTS

We are grateful to David Walsh, Robert A. Cutietta, Frederick Lesemann, and Sharon Lavery for their help in various phases of the experiment and to Rachael Lapidis and Jing Shen for discussions.

Bachem, A. (**1937**). "Various types of absolute pitch," J. Acoust. Soc. Am. **9**, 146–151.
Bachem, A. (**1955**). "Absolute pitch," J. Acoust. Soc. Am. **27**, 1180–1185.
Baharloo, S., Johnston, P. A., Service, S. K., Gitschier, J., and Freimer, N. B. (**1998**). "Absolute pitch: An approach for identification of genetic and nongenetic components," Am. J. Hum. Genet. **62**, 224–231.
Bates, E. (**1992**). "Language development," Curr. Opin. Neurobiol. **2**, 180–185.

Burns, E. M., and Campbell, S. L. (**1994**). "Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: Examples of categorical perception?," J. Acoust. Soc. Am. **96**, 2704–2719.
Curtiss, S., (**1977**). *Genie: A Psycholinguistic Study of a Modern Day "Wild Child"* (Academic, New York).
Dennis, M., and Whitaker, H. A. (**1976**). "Language acquisition following hemidecortication: Linguistic superiority of the left over the right hemisphere," Brain Lang **3**, 404–433.
Deutsch, D. (**2002**). "The puzzle of absolute pitch," Curr. Dir. Psychol. Sci. **11**, 200–204.
Deutsch, D. (**2006**). "The enigma of absolute pitch," Acoustics Today **2**, 11–19.
Deutsch, D., Henthorn, T., and Dolson, M. (**2004**). "Absolute pitch, speech, and tone language: Some experiments and a proposed framework," Music Percept. **21**, 339–356.
Deutsch, D., Henthorn, T., Marvin, E., and Xu, H.-S. (**2006**). "Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period," J. Acoust. Soc. Am. **119**, 719–722.
Doupe, A. J., and Kuhl, P. K. (**1999**). "Birdsong and human speech: Common themes and mechanisms," Annu. Rev. Neurosci. **22**, 567–631.
Duchowny, M., Jayakar, P., Harvey, A. S., Resnick, T., Alvarez, L., Dean, P., and Levin, B. (**1996**). "Language cortex representation: Effects of developmental versus acquired pathology," Ann. Neurol. **40**, 31–38.
Fujisaki, W., and Kashino, M. (**2002**). "The basic hearing abilities of absolute pitch possessors," Acoust. Sci. & Tech. **23**, 77–83.
Gregersen, P. K., Kowalsky, E., Kohn, N., and Marvin, E. W. (**2001**). "Early childhood music education and predisposition to absolute pitch: Teasing apart genes and environment," Am. J. Med. Genet. **98**, 280–282.
Henthorn, T., and Deutsch, D. (**2007**). "Ethnicity versus early environment: Comment on 'Early childhood music education and predisposition to absolute pitch: Teasing apart genes and environment' by Peter K. Gregersen, Elena Kowalsky, Nina Kohn, and Elizabeth West Marvin (2000)," Am. J. Med. Genet. **143A**, 102–103.
Johnson, J. S., and Newport, E. L. (**1989**). "Critical periods in second language learning: The influence of maturational state on the acquisition of English as a second language," Cognit. Psychol. **21**, 60–99.
Kaufman, D., and Aronoff, M. (**1991**). "Morphological disintegration and reconstruction in first language attrition," in *First Language Attrition*, edited by H. W. Seliger and R. M. Vago (Cambridge University Press, Cambridge, UK).
Keenan, J. P., Thangaraj, V., Halpern, A., and Schlaug, G. (**2001**). "Planum temporale and absolute pitch," Neuroimage **14**, 1402–1408.
Kuhl, P. K., Tsao, F.-M., and Liu, H.-M. (**2003**). "Foreign language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning," Proc. Natl. Acad. Sci. U.S.A. **100**, 9096–9101.
Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). "Linguistic experience alters phonetic perception in infants by 6 months of age," Science **255**, 606–608.
Lane, H. L. (**1976**). *The Wild Boy of Aveyron* (Harvard University Press, Cambridge, MA).
Lennenberg, E. H. (**1967**). *Biological Foundations of Language* (Wiley, New York).
Lockhead, G. R., and Byrd, R. (**1981**). "Practically perfect pitch," J. Acoust. Soc. Am. **70**, 387–389.
Miyazaki, K. (**1988**). "Musical pitch identification by absolute pitch possessors," Percept. Psychophys. **44**, 501–512.
Miyazaki, K. (**1992**). "Perception of musical intervals by absolute pitch possessors," Music Percept. **9**, 413–426.
Miyazaki, K. (**2004**). "How well do we understand absolute pitch?," Acoust. Sci. & Tech. **25**, 426–432.
Miyazaki, K., and Rakowski, A. (**2002**). "Recognition of notated melodies by possessors and nonpossessors of absolute pitch," Percept. Psychophys. **64**, 1337–1345.
Newport, E. L. (**1990**). "Maturational constraints on language learning," Cogn. Sci. **14**, 11–28.
Nicoladis, E., and Grabois, H. (**2002**). "Learning English and losing Chinese: A case study of a child adopted from China," Internat. J. Biling. **6**, 441–454.
Profita, J., and Bidder, T. G. (**1988**). "Perfect pitch," Am. J. Med. Genet. **29**, 763–771.
Rakowski, A. (**1972**). "Direct comparison of absolute and relative pitch," in *Symposium on Hearing Theory*, edited by F. A. Bilsen (Instituut voor Perceptie Underzoek, Eindhoven, The Netherlands), pp. 105–108.
Rakowski, A., and Miyazaki, K. (**2007**). "Absolute pitch: Common traits in music and language," Arch. Acoust. **32**, 5–16.
Rakowski, A., and Morawska-Büngeler, M. (**1987**). "In search of the criteria for absolute pitch," Arch. Acoust. **12**, 75–87.
Saffran, J. R., and Griepentrog, G. J. (**2001**). "Absolute pitch in infant auditory learning: Evidence for developmental reorganization," Dev. Psychol. **37**, 74–85.
Schlaug, G., Jaencke, L., Huang, Y., and Steinmetz, H. (**1995**). "In vivo evidence of structural brain asymmetry in musicians," Science **267**, 699–701.
Takeuchi, A. H., and Hulse, S. H. (**1993**). "Absolute pitch," Psychol. Bull. **113**, 345–361.
Ventureyra, V. A. G., Pallier, C., and Yoo, H.-Y. (**2004**). "The loss of first language phonetic perception in adopted Koreans," J. Neurol. **17**, 79–91.
Ward, W. D. (**1999**). "Absolute pitch," in *The Psychology of Music*, 2nd ed., edited by D. Deutsch (Academic, San Diego), pp. 265–298.
Wayland, R. P., and Guion, S. G. (**2004**). "Training English and Chinese listeners to perceive Thai tones: A preliminary report," Lang. Learn. **54**, 681–712.
Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., and Evans, A. C. (**1998**). "Functional anatomy of musical processing in listeners with absolute pitch and relative pitch," Proc. Natl. Acad. Sci. U.S.A. **95**, 3172–3177.

Trombone bore optimization based on input impedance targets

Alistair C. P. Braden,[a] Michael J. Newton, and D. Murray Campbell
School of Physics, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom

(Received 28 August 2008; revised 2 February 2009; accepted 2 February 2009)

Optimization methods based on input impedance target functions have been proposed for the design of brass musical instruments. Criteria for target functions in trombone bore optimization are discussed, drawing on experimental input impedance data from a variety of high-quality trombones of differing sizes. An "inharmonicity plot" is introduced and used to aid the interpretation of impedance curves. An efficient optimization technique is described and is shown to be capable of predicting bore changes which achieve specified modifications to the input impedance curve while maintaining a smoothly-flaring bell contour. Further work is required to clarify the relationship between input impedance targets and the preferences of professional players.
 [DOI: 10.1121/1.3087423]

PACS number(s): 43.75.Fg [NHF] Pages: 2404–2412

I. INTRODUCTION

A player produces a note on a brass musical instrument by expelling a stream of air through the lips, which are pressed against the mouthpiece of the instrument. For an appropriate mechanical configuration, and a suitable range of fluid parameters, the flow of air over the lip surfaces causes them to destabilize and to vibrate. The resulting periodic modulation of the volume flow through the lips acts as an acoustical source, and is the origin of the musical sound radiated by the instrument. When the lip vibration frequency is close to the frequency of one of the acoustical modes of the downstream resonator, a significant pressure variation develops in the mouthpiece, leading to an acoustical feedback from the resonator to the lips. This feedback loop encourages the development of a powerful self-sustained oscillation of the coupled system at a frequency close to that of the acoustical mode in question.

At least for moderate playing levels,[1] the brass instrument may be modeled as a non-linear flow control valve coupling a linear mechanical oscillator (the lips) to a linear acoustical resonator (the air column).[2] The playing behavior of the instrument depends on the properties of both the lips and the resonating tube. In learning to play the instrument, the player develops expertise in lip control; the role of the instrument designer is to optimize the acoustical behavior of the resonating tube when coupled to the lips of a player. The linear acoustical properties of the resonator are described by the input impedance $Z_{\mathrm{in}}=p_{\mathrm{in}}/U_{\mathrm{in}}$, where p_{in} is the acoustic pressure in the mouthpiece and U_{in} the volume flow into the mouthpiece. For a conventional brass instrument such as a trumpet, trombone, or horn, Z_{in} is almost completely determined by the bore-profile $r(x)$, where r is the tube radius and x the axial distance from the mouthpiece entrance plane.

In the 1970s, Smith and Daniell developed a method for calculating the bore perturbations necessary to achieve specified changes in the input impedance curves of brass instruments.[3] More recently, Kausel[4] and Noreland[5] described approaches to the optimization of bore-profiles of brass instruments using targets based on input impedance curves. However, this approach relies on the ability to define the target input impedance curve; it is not yet clear which features of the input impedance curve correspond to optimum behavior as judged by the player.

Poirson *et al.*[6] conducted research into the effect of mouthpiece depth on trumpet timbre, using the spectral centroid to measure the brightness. The results show a correlation (but not necessarily a causal relationship) between the brightness and the inharmonicity, magnitude, and Q-factor of the input impedance peak corresponding to the second harmonic of the played frequency.

Bertsch and co-workers[7–9] made extensive studies into the correlation between input impedance and playing characteristics as determined through blind testing. Chick *et al.*[10] examined the inharmonicity of various French Horns in all valve positions. An extensive comparison of objective and subjective tests of trumpet leadpipes by Poirson *et al.* suggested that the optimum impedance curve as determined by player choices did not correspond to perfect harmonic alignment of acoustic resonances.[6]

Several previous studies[11–17] have identified minor discrepancies between measured and calculated input impedance curves for wind instruments. These discrepancies can arise from uncertainties and calibration problems in the experimental techniques, or from approximations made in the calculations. Real instruments have been successfully optimized in spite of these discrepancies,[18] although they may limit the application of this approach to very fine work. The development of improved optimization techniques is not constrained by this problem, since the efficacy of such techniques as applied to theoretical instruments is independent of the accuracy of the underlying impedance calculations.

In Sec. II of this paper, the use of the input impedance curve as a target for trombone optimization is further explored. Measurements of the input impedance of several trombones which are very highly rated by professional players are presented and analyzed, using a plotting method which allows simultaneous visualization of the resonance

[a]Electronic mail: alistairbraden@gmail.com

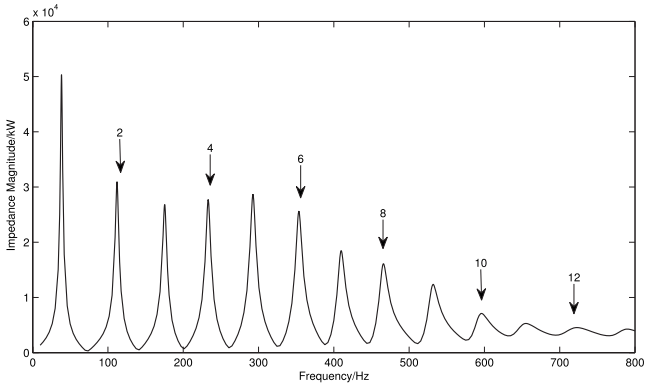


FIG. 1. Theoretical impedance curve of a tenor trombone with slide closed. Peaks comprising the regime of B♭$_2$ highlighted.

frequencies, peak amplitudes, and peak widths. In Sec. III an improved computational method for bore optimization is described, and its use is illustrated in Sec. IV by predicting the bore changes required to achieve specified changes in the frequencies and amplitudes of input impedance peaks of a trombone. Finally, Sec. V discusses the usefulness and limitations of the currently available techniques, and reviews the work which needs to be done to provide a firm foundation for the choice of input impedance targets for brass instrument optimization.

II. INPUT IMPEDANCE TARGETS FOR OPTIMIZATION

A. Measurement and analysis of input impedance curves

A typical input impedance curve for a tenor trombone is illustrated in Fig. 1. Peaks in the curve occur at frequencies corresponding to air column standing waves with a pressure antinode at the mouthpiece. Since the lips behave as a pressure-controlled valve, each impedance peak is close in frequency to one of the pitches which can be strongly sounded; these pitches are known as the natural notes of the instrument.

While an input impedance plot unquestionably contains the essential information about the linear acoustical behavior of an instrument, a number of important features are not obvious when examining such a plot. The most important of these features is probably the mode inharmonicity, which describes the extent to which the peak frequencies deviate from a single harmonic series. Chick *et al.*[10] used the method of *Equivalent Fundamental Pitch* to plot the mode frequencies of horns in a manner which highlights the deviation from perfect harmonicity. For the ith resonance f_i, the fundamental frequency of which it is an exact ith harmonic is calculated, along with the intonation of this equivalent fundamental pitch relative to an arbitrary reference frequency F. This gives a measure, in cents, of the individual harmonic alignment of the peak frequencies, and therefore how closely the peaks collectively match a harmonic series. An example EFP plot is given in Fig. 2; the modes of an instrument with perfect harmonicity would lie on a vertical line on this plot. In this paper, F is chosen such that the fourth resonance lies on the vertical axis.

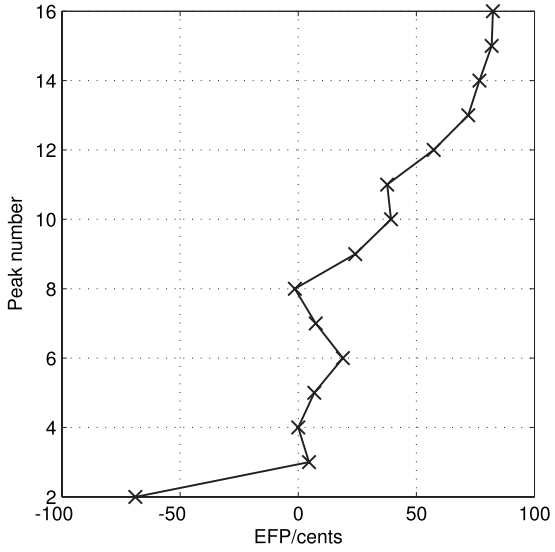


FIG. 2. EFP plot of a typical large-bore tenor trombone, showing resonances 2–16 (F=58.3 Hz).

Although the inharmonicity of the resonances is easily read from an EFP plot, one major disadvantage is the lack of data describing the magnitude and Q-factor of the peaks. The influence of a small, narrow, harmonic peak may be somewhat less than that of a tall, broad inharmonic peak; the EFP does not represent this and can therefore be misleading. For this purpose, an *inharmonicity plot* is introduced; this is a development of the EFP plot including both the magnitude and the Q-factor. For each mode, the impedance magnitude at a series of frequencies corresponding to a pitch interval ±100 cents around the reference frequency F is plotted on an EFP plot as a thin horizontal band, with color or grayscale denoting the impedance magnitude. Figure 3 shows the inharmonicity plot of a typical tenor trombone (this, along with the EFP plot in Fig. 2, is taken from the impedance curve in Fig. 1). The inharmonicity, relative magnitudes, and shapes of the peaks can be viewed simultaneously on this plot; an instrument with strong narrow and harmonic impedance peaks would appear as a narrow white vertical band.

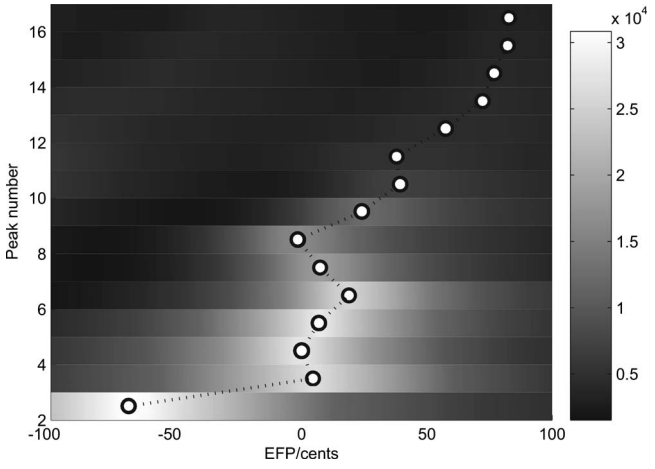


FIG. 3. Inharmonicity plot of a typical large-bore tenor trombone, showing resonances 2–16 (F=58.3 Hz).

B. Inharmonicity and mode-locking

Backus and Hundley[19] first suggested that an approximately harmonic alignment of the resonator peaks would help to maintain stable oscillations of the lips. For example, B♭$_2$ played on a trombone excites the resonances numbered 2, 4, 6, etc., represented by the corresponding input impedance peaks (Fig. 1). Fletcher[20] pointed out that, if the inharmonicity of the modes were sufficiently small, a self-sustained vibration regime could be established in which several modes maintained a fixed phase relationship. This mode-locked regime was described by Benade[21] as a "cooperative regime of oscillation."

Because of the non-linear character of the lip valve, the relative influence of each of the higher resonances on the regime is dependent on the dynamic level played; qualitatively, at pianissimo the excitation is nearly sinusoidal and only the sounded resonance has a significant influence, whereas at forte (excluding "brassy" cases[22]) the higher resonances have a much larger effect. It should be noted that some peaks (e.g., 4, 6, and 8) are part of the regimes of several sounded notes, whereas others (those of prime index) contribute only to one regime.

Worman[23] attempted to describe the overall intonation and resonance characteristics of a regime based on the sum-function

$$\mathrm{SF}(f) = \frac{1}{n}\sum_{n} \mathrm{Re}(Z_0(nf)), \tag{1}$$

where n is the harmonic number, Z_0 the input impedance, and f the frequency in question. Under this assumption, a perfectly-harmonic set of resonances would provide the tallest possible peaks in the sum-function, and therefore offer the strongest response to the player. The harmonic content in the radiated sound would be maximized (for the given peak magnitudes) and the intonation of the played regimes would have integer relationships. A modified version of the sum-function, altering the weighting such that higher peaks have a progressively greater influence at higher dynamics, was originally proposed by Worman and has recently been implemented in commercial software.[24]

The most strongly coupled oscillation regime would occur for a perfectly-harmonic series of acoustical resonances. However, it is not obvious that this should be the goal of the brass instrument designer, who must balance various issues including timbre, playability, stability, and intonation. A perfectly-harmonic series of bore resonances may produce a very stable oscillation, provided the player can match the embouchure to the note, but the resulting regime would be very difficult to "bend" or "lip" in musically expressive ways that alter the playing frequency of the note. Jazz players, and performers on natural instruments like the baroque trumpet, often require to bend a note away from its nominal pitch, a technique that is certainly easier when the instrument bore has some inharmonicity.[25]

The structure of musical scales also implies that a perfectly-harmonic series of resonances would not necessarily yield perfect musical intonation in all circumstances. Apart from the octave, none of the intervals of the equally-tempered scale corresponds exactly to an interval of the harmonic series; playing in tune requires frequent subtle modifications in playing pitch to fit the musical context. The balance to be drawn between stability and flexibility of intonation will depend on the playing style and repertoire of the individual musician.

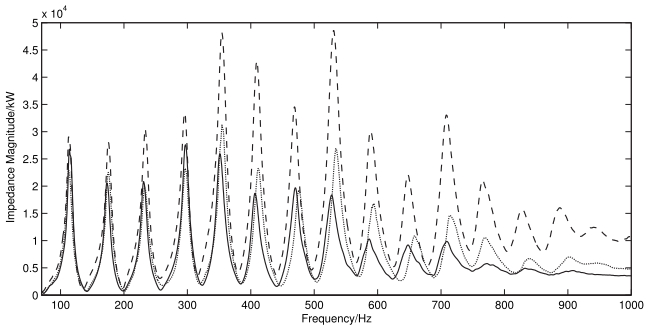


FIG. 4. Measured impedance plots for trombones A (dashed), B (dotted), and C (solid).

C. Input impedance measurements on high-quality trombones

In choosing an instrument, a professional player takes many different criteria into account.[6,8] To investigate how these choices are reflected in the input impedance curves, impedance measurements were made on three contrasting professional-standard models of trombone, all from the same highly-regarded British manufacturer. Trombone A was a medium-bore jazz trombone, B was a large-bore orchestral tenor, and C a bass trombone; each was fitted with an appropriate mouthpiece. The measurements were made using the BIAS input impedance measurement system.[24]

Figure 4 shows the input impedance magnitude plots. The most striking difference is in the peak heights for trombone A, which are consistently taller than for the larger bore instruments. The peak heights of the bass trombone C are generally, but not always, lower than those for B, particularly in the upper range of the instrument. The peaks of A in the upper range are particularly tall, which may be linked with the comparative ease of playing in this register with this instrument.

From these magnitude plots are produced inharmonicity plots (Fig. 5). It is clear that the peak frequencies do not represent an exact harmonic series. For each trombone, the first mode is several semitones below the nominal pitch of B$_1^{\flat}$ and is not shown on the EFP plot. The second peak is also rather flat, although markedly less so for the bass trombone C than for the tenor trombones. Given the similar peak heights and shapes in the lower ranges of these instruments, and the knowledge that the bass is generally preferred to the tenor for low pitch music, this preference can be tentatively linked with the large difference in the alignment of mode 2.

Each instrument shows a general right-hand diagonal trend on the EFP plot, denoting that higher peaks are tuned to a progressively sharper equivalent fundamental; the effect is considerably lesser in the smaller trombone. Similar results were taken with the same BIAS equipment on French Horns,[10] and demonstrate that B♭ horns have peaks which

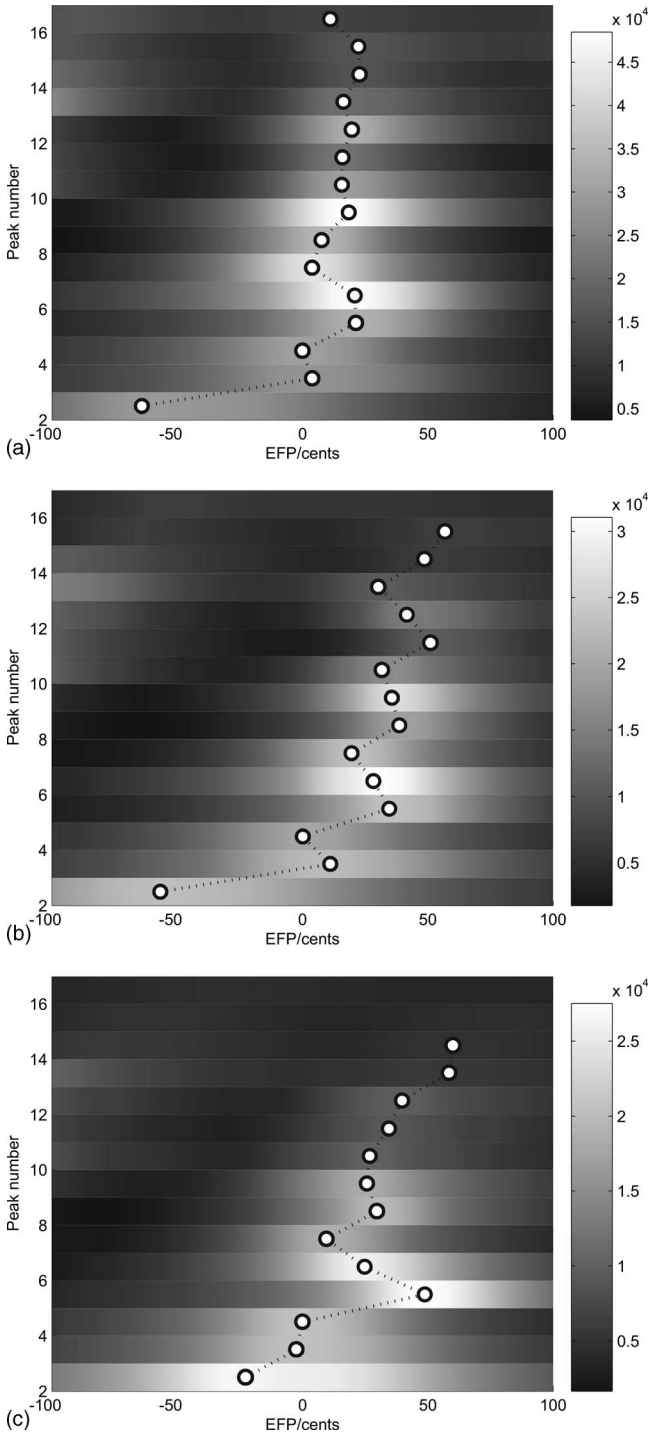


FIG. 5. Inharmonicity plots, showing resonances 2–16.

are considerably closer to a harmonic series than those of the trombones here. From these two studies, given that these instruments are considered of first-rate quality, it can be concluded that inharmonicity of this kind is a desirable feature contributing to the characteristic sound of the trombone, and particularly the large-bore trombone.

III. BORE OPTIMIZATION USING SPECIFIED INPUT IMPEDANCE TARGETS

For some years now, computer optimization techniques have been applied to brass instrument design problems. By taking an established model to calculate the impedance of an instrument with given bore-profile, and combining it with a search algorithm, a computer can be used to find the bore-profile whose impedance best matches specified performance characteristics (such as the location of resonances). Such optimization techniques allow many possible instrument designs to be tried quickly, without the material expenses of building many prototypes. Optimization can be used for *bore reconstruction* (deducing an unknown bore from a known impedance) and for *performance optimization* (modifying an existing design to alter certain characteristics while maintaining others).

The first attempt at such optimization was by Kausel,[4] representing the instrument using a piece-wise linear interpolation of points along the bore ("point-wise") and optimizing with the 0th-order Rosenbrock search algorithm. This method is commercially available and has been used by a manufacturer to design trumpet leadpipes.[18] Noreland[5,26] used a point-wise representation of the second derivative of the bore radii and the gradient-based Levenberg–Marquardt algorithm to give smoother bore-profiles. Both of these techniques assume plane-wave propagation. Similar techniques have been applied to the optimization of xylophones[27,28] and to leak detection.[29]

This section describes briefly the computer optimization technique developed by Braden, further details of which are given elsewhere.[15,30] This technique combines n-mode impedance modeling with the Rosenbrock optimization algorithm to search within a space of possible instrument designs to find that one whose input impedance best matches target criteria. The technique given is well-suited to optimizing complete instruments with realistic geometries, giving smooth bell flares. We show here results demonstrating the latter, using a design space constrained to include only "reasonable trombone shapes."

Any design optimization problem must combine a numerical representation of the object being designed, a quantitative method of evaluating possible designs, and an algorithm for the next "try" given previous tries.

A. Objective function

Input impedance magnitude is an objective acoustic property, contains much useful information about the performance of the instrument, is experimentally measurable, and can be calculated numerically; this combination of features allows us to center our optimization around it. The peak locations (frequency and magnitude) have the greatest influence on the performance, though other properties, such as the width and shape of peaks, also have an effect. The aim is to find the instrument shape whose impedance properties best match certain targets; these targets will be defined in terms of another impedance curve and/or certain specific characteristics of the impedance peaks.

Given an instrument shape, which is denoted symbolically by α, a method for computing the input impedance must be chosen. Previously, this has been a plane- or spherical-wave model,[31] which both Kausel and Noreland employed; the results in this paper were also computed using

this method. However, the optimization technique described here has also been successfully employed[15] with a model including higher modes.[12,14,32] Qualitatively similar optimization results were found with either model (with appropriate targets), so either model could have been chosen in principle. In practice, the computationally rather less expensive plane-wave model is more convenient for repeated experiments investigating the full capabilities of the optimizer, as is the primary goal of this paper, and was therefore chosen.

An objective function must be chosen to compare the impedance properties of each "tested" instrument to target impedance data. Three objective functions are used here in different combinations depending on the task, each giving scores between 0 (a perfect match) and 1 (no match). Typical tuning-parameter values are given. The objective functions are formulated as a minimization problem. The first function is a windowed linear least-squares comparison of the entire (discrete) impedance magnitude curve with N_{pt} points:

$$\mathcal{O}_1(\alpha) = \frac{1}{N_{\mathrm{pt}}} \sum_{i=1}^{N_{\mathrm{pt}}} \begin{cases} \dfrac{(\delta z_i)^2}{\nu^2} & \text{if } \delta z_i < \nu, \\ 1 & \text{if } \delta z_i \geqslant \nu, \end{cases} \tag{2}$$

where $\delta z_i = |z(f_i) - \bar{z}(f_i)|$, i.e., the difference between tested impedance magnitude z and target $\bar{z}$ measured at a series of frequencies f_i, and ν is the half-width of the window (usually 100–300 kΩ). Any valid method for computing $z(f_i)$ from α may be used. Both Kausel and Noreland used this least-squares approach but without the windowing, which eliminates cases where a very bad feature can overwhelm the more subtle difference in score between another feature being good and very good, thus improving overall convergence. In addition to the least-squares approach, a windowed-Gaussian comparison is used, scoring only the peak frequencies:

$$\mathcal{O}_2(\alpha) = \frac{1}{N_{\mathrm{pk}}} \sum_{i=1}^{N_{\mathrm{pk}}} \begin{cases} \dfrac{1 - \exp\left(\dfrac{-\mu_\phi (\delta\phi_i)^2}{\nu_\phi^2}\right)}{1 - \exp(-\mu_\phi)} & \text{if } \delta\phi_i < \nu_\phi, \\ 1 & \text{if } \delta\phi_i \geqslant \nu_\phi, \end{cases} \tag{3}$$

where $\delta\phi_i = |\phi_i - \bar{\phi}_i|$, i.e., the difference between tested peak-frequency ϕ_i and target peak-frequency $\bar{\phi}_i$ of peak i, the window half-width ν_ϕ (typically 10 Hz), μ_ϕ is a "strictness" parameter (typically 20), and N_{pk} the number of impedance peaks being tested. The function joins smoothly at the window bounds $\phi_i \pm \nu_\phi$ to take a value of 1. The gradient of the Gaussian function is much steeper at a moderate distance from the target than the equivalent least-squares function, and consequently offers improved convergence speed in this region.

A very similar function $\mathcal{O}_3$ is also used for the peak heights, using difference $|z(\phi_i) - \bar{z}(\bar{\phi}_i)|$ and magnitude-dependent window half-width $\nu_z z(\phi_i)$, with parameters ν_z and μ_z (typically 0.05 and 10, respectively). An overall objective function is then defined as

$$\mathcal{O}(\alpha) = \frac{\zeta_1 \mathcal{O}_1(\alpha) + \zeta_2 \mathcal{O}_2(\alpha) + \zeta_3 \mathcal{O}_3(\alpha)}{\zeta_1 + \zeta_2 + \zeta_3}, \tag{4}$$

where $\zeta_{1,\ldots,3}$ are weights which can take any real value, but are usually 0 or 1. The goal is to find the α which minimizes $\mathcal{O}(\alpha)$.

B. Instrument representation

The convergence rate of optimization algorithms is strongly dependent on the number of design variables N_v and the size of the design space; by reducing each of these, the performance of the optimization can be improved. Care must be taken to avoid pre-determining the eventual solution by constraining the space too restrictively; the optimizer must be given enough freedom to explore many reasonable solutions.

In Kausel's method, no attempt is made to constrain inappropriate designs, so many outlandish shapes are possible; for example, Fig. 6 in Ref. 33 is too "jagged" to be a realistic instrument. Noreland's smoothing method makes jagged results unlikely, but instead "wiggly" shapes are common (for example, Fig. 14 in Ref. 26). Designs such as these would be rejected out of hand by any manufacturer, regardless of their playing characteristics; the present method attempts to remove them from the design space altogether.

To a first approximation, a trombone is a long cylindrical tube, of given length and radius, connected smoothly to a Bessel-horn, defined as

$$r(x) = b(-x)^{-\gamma}, \tag{5}$$

where r is the bore radius a distance x along the axis of instrument, γ is a flare constant, and b is a constant defined in terms of specified length and input/output radii. Given appropriate constraints, it is possible to define a five-dimensional space (i.e., $N_v = 5$) of possible designs, all of which are reasonable (albeit simple) trombone shapes. Bessel-horns, however, are not exact matches for real trombone bells, so the design space here is too limited. The bell is therefore described as a number of shorter Bessel-horns placed end-on-end,[30] to achieve much closer *piece-wise Bessel-horn interpolations* of real instruments. By confining the design space in this manner, many unacceptable possible solutions have been eliminated, significantly reducing the size of the design space. As with other piecewise interpolation techniques, a small discontinuity of gradient at the joints between pieces must be accepted; some unrealistic designs with large gradient discontinuities therefore exist in the space, but are proportionately far less frequent than for other representations.

In practice, the shape of a real trombone bell can be very closely described by five Bessel-horns of different flares, giving $N_v = 12$, where one variable is the length of the sections, six describe the input and output radii of the sections, and five give their flare coefficients (previous approaches typically require $N_v \sim 100$). Use of such a higher-order parametrization of the design of the instrument therefore con-

fers a significant reduction in the number of design variables as well as a reduction in the size of the design space, giving an increase in optimization speed.

Other practical necessities in realistic instrument design, such as discontinuities caused by the joins between sections, tuning slides, etc., can be included in such parametrizations. In this manner, a description can be given of the complete general shape of the instrument to be designed;[30] this is termed a *design template*, and it is a representation of the detailed geometry of the instrument without the exact dimensions. A template can be constructed that can describe all trombones fairly closely; however, in practice it is more effective to have more detailed templates for more specific problems, such as tenor and bass trombones. Preset mouthpiece geometries can be specified and held fixed throughout the optimization.

Approaching optimization problems in this manner trades off generality against improved convergence. One of the advantages of the pointwise approach was the suitability to any given problem with little or no prior knowledge of the instrument shape—so-called "black-box" problems. The approach given here assumes that the rough shape of the instrument is already known, while this is less attractive in principle, the disadvantage is negligible in practice. Given that the primary objectives are the reconstruction and optimization of brass instruments, it is not unreasonable to assume that the class of the instrument (e.g., trumpet or trombone) is already known, and therefore that the rough shape is also known; a template can thus be chosen from prior knowledge of these instruments. The constraints are set to be large enough that the design space includes all known instruments of that class, so that it is still general enough to include all solutions to the specific problem in hand, but removes many of the unreasonable solutions that another, more general, approach would include. This approach favors pragmatism over rigorous generality, but it should be noted that templates can be constructed to consider general problems in the same (point-wise) way as Kausel,[4] so, rather than being lost, this generality is merely put aside unless needed.

C. Optimization algorithm

Noreland successfully applied a gradient-based method to instrument optimization. Use of such methods, however, requires a derivative of the objective function with respect to the design variables, and therefore a derivative of the input impedance. Noreland provided such a derivative for the plane-wave model. The technique described here was developed to be applicable to a broader range of impedance techniques, including methods for higher modes[14] and bent waveguides.[34] In principle, derivatives of these methods can be found and combined with a gradient-based method. Such derivation is outside the scope of the current work and has not yet been attempted; the present study is therefore restricted to the use of optimization algorithms which do not require derivatives. In the vein of Kausel, the Rosenbrock[35] method is chosen; this has been found to be superior to evolutionary approaches even in low-N_v parametrized problems.[30]

IV. OPTIMIZATION RESULTS

A. Tuning peak 2

It has been established through experimental comparison of a tenor and bass trombone that the tuning of peak 2 relative to the higher even-numbered peaks (particularly peak 4) has a significant effect on the lower playing register of the instrument. In this experiment, an attempt was made to improve a popular design of tenor trombone by shifting peak 2 into a closer alignment without compromising any other resonance properties.

The starting point was based on detailed geometrical measurements of a large-bore tenor trombone. The impedance of the instrument was then calculated with the spherical-wave model and the peak frequencies set as an optimization target, with the following modifications: the target frequency of peak 2 was shifted from 112.0 to 115.0 Hz (a shift of +45 cents), and the target frequency of peak 5 was shifted from 292.5 to 295.0 Hz (a shift of +15 cents). It was found that allowing freedom in the location of peak 5 granted greater flexibility in the location of peak 2.

Figure 6 shows that the optimizer produced a solution which closely matches the target set. Peaks 2 and 5 are now located at 114.7 and 295.1 Hz, respectively. Peaks 3 and 4 have not been significantly altered. It is evident from the EFP plot [Fig. 6(b)] that the desired re-alignment of peak 2 has occurred. As a compromise, the higher peaks have all been shifted by between +5 cents and +20 cents; these are all rather smaller than the +41 cents shift of peak 2. The impedance magnitudes of the peaks have been affected, notably peaks 3–6, which have been shortened, and peaks 8 and 9, which are taller.

Figure 6(c) shows the optimized bore-profile. The bore design is smooth, realistic, and could be built by a manufacturer with no additional difficulty; it can be described as a modified version of an existing design. The bore of the cylindrical section (i.e., the main slide) has been reduced from 6.95 mm radius (or, in the conventions of the instrument industry, 0.547 in. diameter) to 6.63 mm radius (0.522 in. diameter). The overall instrument length is some 26 mm shorter than before, and the bell contour has been subtly altered, as has the taper of the tuning slide. Given the performance target, these results are somewhat surprising; since the optimization is attempting to replicate a feature of a bass trombone, it might have been expected that the bore of the cylindrical section would have increased.

Clearly the trombone design space is complicated. Certain regions of high performance are well-established in the industry, and it may be that the optimizer has uncovered a region which might have otherwise been unexplored. Musical judgments of the success of the optimization will not be available until an example has been built and tested; however, the exercise has demonstrated the capability of the software to perform whole-instrument intonation optimization.

B. Modifying peak magnitudes

In a second test of the optimization software, an attempt was made to modify the magnitudes of specified impedance peaks. The target impedance was again a modification of the

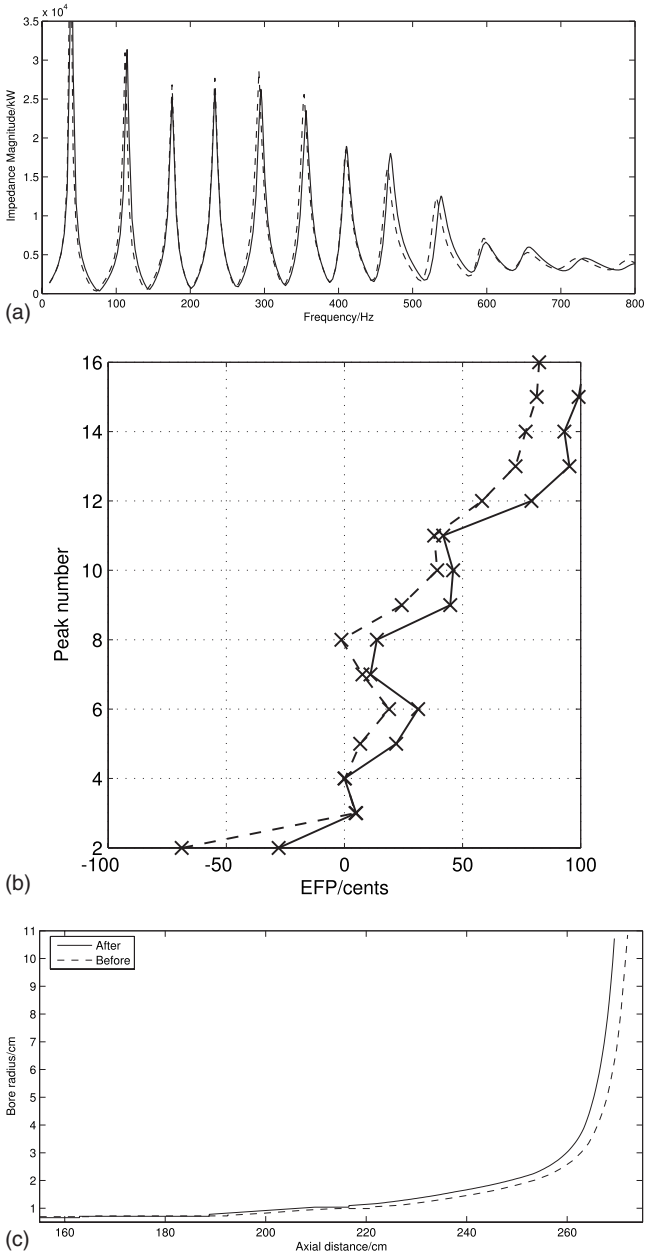


FIG. 6. Results of tuning peak 2 before (dashed) and after (solid) optimization.

measured impedance curve for the large-bore tenor trombone: this time peaks 8, 9, and 10 were targeted to increase in magnitude by 10%, and no target was set for the magnitude of peaks 1–7. The peak frequencies were left unchanged.

It can be seen from Fig. 7(a) that the optimization was successful, increasing peak 8 from 16 110 to 17 602 kΩ, peak 9 from 12 358 to 13 668 kΩ, and peak 10 from 7094 to 8064 kΩ, respectively, increases of 9.2%, 10.6%, and 13.6%. Certain lower-index peaks were reduced in amplitude, most notably peak 5 from 28 671 to 25 820 kΩ, a decrease of 9.9%. The frequency (and therefore EFP) of peaks 2–10 was not significantly changed; higher peaks were each made roughly 5 cents sharper. The bore-profile [Fig. 7(b)] again shows a reduction in the radius of the main slide from 6.95 mm (0.547 in. diameter) to 6.65 mm (0.524 in.

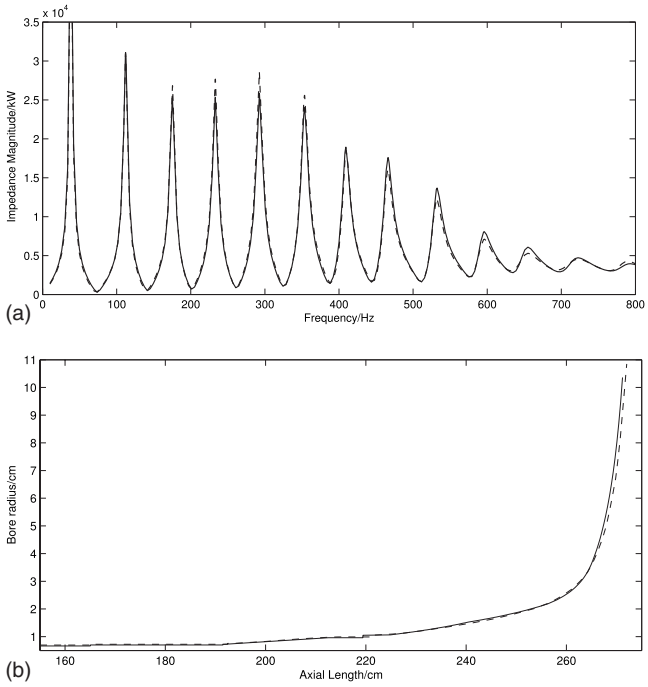


FIG. 7. Results of increasing the magnitudes of peaks 8–10 before (dashed) and after (solid) optimization.

diameter), and some subtle changes (a maximum of 0.5 mm) to the bell contour. The experiment was repeated but with 20% increases specified; the optimizer was unable to converge to a geometrically-satisfying solution, suggesting that such a result was not possible within the constraints of the design space.

These results demonstrate that it is possible to modify the impedance envelope without substantially modifying the peak intonation. The changes made here to the envelope are of similar proportion to those found in the previous experiment, but had only a small effect on intonation. As a result of these changes, the optimized trombone would be expected to produce a somewhat brighter timbre than its predecessor and possess a higher register which is somewhat easier to play.

V. CONCLUSIONS

Modern brass instruments have sets of acoustic resonances which approximate closely to a harmonic series. When a normal note is sounded, the player's lips interact with a subset of these resonances to create a mode-locked regime of oscillation. The pitch of the radiated sound is that of the lip vibration, which will not necessarily correspond exactly to one of the acoustic resonance frequencies. The magnitude of the impedance at the sounding frequency and at its harmonics will affect the stability of the note and the timbre of the sound.

Is it a reasonable assumption that the peak frequencies of an "ideal" brass instrument should be harmonically related? Experimental measurements of several professional-standard models of trombone, intended for use in orchestral and in jazz idioms, suggest that this is not the case: having higher resonances tuned to a progressively sharper equivalent fundamental appears to be a desirable characteristic of

trombones, particularly those of larger bore. From a purely musical standpoint, the designer's objective is not to maximize the stability or brightness of tones, but to produce an instrument which is easy to play, sufficiently in tune, and produces a timbre which is appropriate to the musical style and context of the performance.

Why might the inharmonicity present in a large-bore trombone be desirable? Consider a hypothetical trombone with many tall harmonic peaks (similar to trombone A in many respects). This instrument would be expected to have a large tonal range and produce a relatively bright timbre.[6,21] If this brightness were undesirable (as it generally is in orchestral settings), then a designer might combat this by reducing the heights of the middle and upper peaks, thus reducing the harmonic content in the sound. However, this would have the knock-on effect of reducing the stability of the higher tones. In order, therefore, to reduce the brightness without affecting the tonal range, some inharmonicity is necessary to reduce the influence of the higher peaks on the lower tones. In reality the orchestral tenor trombone may be thought of as a compromise between these two solutions, featuring shorter, less harmonic peaks than the jazz trombone, but having a similar tonal range. Trombonists are accustomed to making small corrections in the position of the hand-slide in any given musical situation for the purposes of fine intonation, and are therefore likely to be more tolerant of small alterations to the intonation of the instrument than would the player of a valved instrument.

The study described here has demonstrated the ability of an improved optimization procedure to propose technologically feasible and visually acceptable brass instrument bore shapes whose impedance curves correspond to a predetermined target. When the target requirements are compatible with the limitations imposed on the bore by non-acoustic design considerations, the optimization converges rapidly to a solution: roughly 1500 designs of the complete instrument are needed to converge to within a design tolerance of 0.01 mm. Lack of convergence indicates that the requirements of the target cannot be met without relaxing the limitations imposed on the design space.

Discrepancies between calculated and experimentally measured input impedance curves[12,15] currently limit the applicability of computer optimization to the solution of subtle problems in real instrument design. The optimization method described here is applicable to any impedance model, and any future improvements in numerical modeling could be immediately applied to optimization problems. This provides a strong motivation to develop further the underlying theoretical models and experimental impedance measurement techniques.

It should also be recognized that the input impedance describes only the linear behavior of the acoustic resonator. Since the interior acoustic pressure in a brass instrument can reach 180 dB in loud playing, non-linear sound propagation plays an important role in the development of timbre with increasing dynamic level. Once the bore-profile has been determined by the optimizer, non-linear effects can be calculated and their significance estimated.

The major limitation in the optimization of brass instrument design remains an inadequate understanding of the ideal target impedance function. The work of Poirson[6] on trumpets has shown the feasibility of a multidimensional scaling approach to finding correlations between player judgments of instruments and objective impedance criteria. Similar studies on a wide range of players and instrument types are required, in order to establish clearly which features of the impedance curves correlate closely with player ratings based on musical judgments. With this information available, the optimizer can become a valuable tool in the production of brass instruments of the highest quality.

[1] A. Hirschberg, J. Gilbert, R. Msallam, and A. P. J. Wijnands, "Shock waves in trombones," J. Acoust. Soc. Am. **99**, 1754–1758 (1996).
[2] D. M. Campbell, "Nonlinear dynamics of musical reed and brass instruments," Contemp. Phys. **40**, 415–431 (1999).
[3] R. A. Smith and G. J. Daniell, "Systematic approach to the correction of intonation in wind instruments," Nature (London) **262**, 761–765 (1976).
[4] W. Kausel, "Optimization of brasswind instruments and its application in bore reconstruction," J. New Music Res. **30**, 69–82 (2001).
[5] D. H. Noreland, "Gradient-based optimisation of brass instruments," in Proceedings of the Stockholm Music Acoustics Conference SMAC03, Sweden (2003).
[6] E. Poirson, J. F. Petiot, and J. Gilbert, "Study of the brightness of trumpet tones," J. Acoust. Soc. Am. **118**, 2656-2666 (2005).
[7] M. Bertsch, "Bridging instrument control aspects of brass instruments with physics-based parameters," in Proceedings of the Stockholm Music Acoustics Conference SMAC03 (2003).
[8] M. Bertsch, K. Waldherr, and W. Kausel, "Sensory evaluation testing of trumpets and correlation with acoustic measurements," in Proceedings of Forum Acusticum, Budapest (2005), pp. 575–580.
[9] M. Bertsch, K. Waldherr, M. Horvat, and K. Jambrošic, "Reliability of trumpet players' judgements during instrument blind-test," in Proceedings of Second Congress of Alps-Adria Acoustics Association and First Congress of Acoustical Society of Croatia. Zagreb, Croatia (2005).
[10] J. P. Chick, C. Lumb, and D. M. Campbell, "Passive acoustic characteristics and intonation problems of modern orchestral horns," in Proceedings of ISMA2004, Nara, Japan (2004).
[11] R. Causse, J. Kergomard, and X. Lurton, "Input impedance of brass musical instruments—Comparison between experiment and numerical models," J. Acoust. Soc. Am. **75**, 241–254 (1984).
[12] N. Amir, V. Pagneux, and J. Kergomard, "A study of wave propagation in varying cross-section waveguides by modal decomposition. Part II. Results," J. Acoust. Soc. Am. **101**, 2504–2517 (1996).
[13] J. P. Dalmont, "Acoustic impedance measurement. Part 1: A review," J. Sound Vib. **243**, 427–439 (2001).
[14] J. A. Kemp, "Theoretical and experimental study of wave propagation in brass musical instruments," Ph.D. thesis, University of Edinburgh (2002).
[15] A. C. P. Braden, "Bore optimisation and impedance modelling of brass musical instruments," Ph.D. thesis, University of Edinburgh (2006).
[16] P. Dickens, J. Smith, and J. Wolfe, "Improved precision in measurements of acoustic impedance spectra using resonance free calibration loads and controlled error distribution," J. Acoust. Soc. Am. **121**, 1471–1481 (2007).
[17] C. J. Nederveen and J. P. Dalmont, "Corrections to the plane-wave approximation in rapidly flaring horns," Acta. Acust. Acust. **94**, 461–473 (2008).
[18] W. Kausel and R. Egger, "The brass-wind-instrument-optimizer as a tool for instrument makers: A case study," in Proceedings of Vienna Talk on the Future of Wind Instruments (2005).
[19] J. Backus and T. C. Hundley, "Harmonic generation in the trumpet," J. Acoust. Soc. Am. **49**, 509–519 (1971).
[20] N. H. Fletcher, "Mode locking in nonlinearly excited inharmonic musical oscillators," J. Acoust. Soc. Am. **64**, 1566–1569 (1978).
[21] A. H. Benade, *Fundamentals of Musical Acoustics* (Oxford University Press, New York, 1976) pp. 402–405.
[22] M. W. Thompson and W. J. Strong, "Inclusion of wave steepening in a frequency-domain model of trombone sound production," J. Acoust. Soc. Am. **110**(1), 556–562 (2001).
[23] W. E. Worman, "Self-sustained non-linear oscillations of medium amplitude in clarinet-like systems," Ph.D. thesis, Case Western Reserve Univer-

sity, Ann Arbor (1971).
[24]W. Kausel, "Brass instrument analysis system," http://www.bias.at (Last viewed January, 2009).
[25]D. Smithers, K. Wogram, and J. Bowsher, "Playing the baroque trumpet," Sci. Am. **254**(4), 104–111 (1986).
[26]D. Noreland, "A gradient based optimisation algorithm for the design of brass-wind instruments," Ph.D. thesis, Uppsala University (2003).
[27]J. Petrolito and K. A. Legge, "Designing musical structures using a constrained optimization approach," J. Acoust. Soc. Am. **117**, 384–390 (2005).
[28]L. Henrique, O. Inácio, J. Paulino, and J. Antunes, "Optimization of vibratory and acoustical components of percussion instruments: Theoretical and experimental results," in Proceedings of Forum Acusticum, Budapest (2005), pp. 485–490.
[29]V. Chilekwa and D. Sharp, "Detection, location and sizing of multiple leaks in a duct," in Proceedings of Forum Acusticum, Budapest (2005), pp. 2569–2572.
[30]A. C. P. Braden, "Optimisation techniques for solving design problems in modern trombones," in Proceedings of Forum Acusticum, Budapest (2005), pp. 557–562.
[31]D. Mapes-Riordan, "Horn modeling with conical and cylindrical transmission-line elements," J. Audio Eng. Soc. **41**, 471–483 (1993).
[32]V. Pagneux, N. Amir, and J. Kergomard, "A study of wave propagation in varying cross-section waveguides by modal decomposition. Part I. Theory and validation," J. Acoust. Soc. Am. **100**, 2034–2048 (1996).
[33]W. Kausel, "Computer optimization of brass wind instruments," in Diderot Forum on Mathematics and Music, Vienna (1999), pp. 227–242.
[34]A. C. P. Braden, "Input impedance of bent waveguides," in Proceedings of the Institute of Acoustics Spring Conference, Southampton (2006), Vol. **28**(1).
[35]H. H. Rosenbrock, "An automatic method for finding the greatest or least value of a function," Comput. J. **3**, 175–184 (1960).

A bulk modulus dependent linear model for acoustical imaging

Jean Martial Mari[a)]
Department of Bioengineering, Royal School of Mines, Imperial College London, South Kensington Campus, London SW7 2BP, United Kingdom

Thierry Blu
Department of Electronic Engineering, Chinese University of Hong Kong, Shatin New Territories, Hong Kong

Olivier Bou Matar
Laboratoire Européen Associé en MagnétoAcoustique de la Matière Condensée (IEMN-DOAE UMR CNRS 8520), Ecole Centrale de Lille, BP 48, 59651 Villeneuve d'Ascq Cedex, France

Michael Unser
Biomedical Imaging Group, EPFL/STI/IMT/LIB, CH-1015 Lausanne VD, Switzerland

Christian Cachard
CREATIS, Université de Lyon, INSA-Lyon, Université Lyon 1, CNRS UMR5220, INSERM U630 Bâtiment Blaise Pascal, INSA, 7 Avenue Jean Capelle, F69621 Villeurbanne Cedex, France

(Received 26 February 2007; revised 24 September 2008; accepted 3 February 2009)

Modeling the acoustical process of soft biological tissue imaging and understanding the consequences of the approximations required by such modeling are key steps for accurately simulating ultrasonic scanning as well as estimating the scattering coefficient of the imaged matter. In this document, a linear solution to the inhomogeneous ultrasonic wave equation is proposed. The classical assumptions required for linearization are applied; however, no approximation is made in the mathematical development regarding density and speed of sound. This leads to an expression of the scattering term that establishes a correspondence between the signal measured by an ultrasound transducer and an intrinsic mechanical property of the imaged tissues. This expression shows that considering the scattering as a function of small variations in the density and speed of sound around their mean values along with classical assumptions in this domain is equivalent to associating the acoustical acquisition with a measure of the relative longitudinal bulk modulus. Comparison of the model proposed to Jensen's earlier model shows that it is also appropriate to perform accurate simulations of the acoustical imaging process.
 [DOI: 10.1121/1.3087427]

PACS number(s): 43.80.Cs, 43.20.Bi, 43.35.Bf, 43.80.Vj [LLT] Pages: 2413–2419

I. INTRODUCTION

Ultrasonic imaging is a complex process during which a focused beam is transmitted through a soft biological tissue and scattered by inhomogeneities stemming from local variations in the intrinsic mechanical properties. Ideally, knowing the electromechanical transfer function of the probe and the measured backscattered signal, called the radiofrequency (RF) signal, should solve the inverse problem, thus retrieving the real scattering profiles. This consists in removing the influence of the acquisition tool and displaying an ultrasonic image of higher quality. Unfortunately, the problems that arise from the complexity of the scattering process[1–6] in the tissues and from the technical limits of the ultrasonic probes render the mathematical inversion an open problem. To be solved, the problem has to be regularized by formulating assumptions on the propagating ultrasound wave, the behavior of the continuous collection of scatterers, and finally the range of variation of the local mechanical properties. Among the possible approaches and the hypothesis these approaches imply, one solution consists in searching for a linear description of the problem,[7–9] transforming the inversion step into a well-known deconvolution process.[10] A linear description means that the response of a collection of scatterers is the sum of the responses of each constituent scatterer. This form then makes it possible to quantitatively solve the inverse problem, while the mathematical expression of the problem can reveal the dependence of the estimated scattering coefficient on a particular intrinsic mechanical parameter. Specifically, this sheds light on the consequences of the assumptions on the nature of the resulting scattering coefficient and also reveals *a posteriori* the approximations required, as a cost for obtaining a simplified description of the acoustical process.

Linear descriptions for the ultrasonic imaging process of soft tissues are well known and have been extensively developed in literature, such as in Refs. 1, 2, 7, and 8. As previously shown, one of their main results consists in expressing the scattering coefficient as a function of the local intrinsic properties of the medium. In the past, Gore and Leeman[1] and

[a)]Electronic mail: jm.mari@imperial.ac.uk

more recently Zemp *et al.*[7] showed that under various assumptions in the field and the insonified medium, the scattering coefficient depends on the adiabatic compressibility coefficient and density. The approach of Zemp *et al.*[7] contributes an original solution to space-varying point spread functions (PSFs) and allows for a statistical approach to the problem. Fatemi and Kak[2] expressed the scattering coefficient as a function of sound speed only, though strong hypotheses were imposed. Jensen[8] derived the scattering coefficient as a function of sound speed and density. This model is very interesting because it provides an expression of the model where each part can be measured or defined, so that one can, respectively, partly retrieve the scatterer distribution or simulate medical ultrasound imaging. Unfortunately, the first-order expression of Jensen's scattering coefficient[8] does not take advantage of the stronger hypothesis involved and the final expression contains unused approximations.

In this paper, it is shown that those expressions of the ultrasonic process and their scattering coefficients, a function of sound speed and density, can be linked to the variations in the sole longitudinal bulk modulus. As a result, the ultrasound scanning process may be interpreted under the classical assumptions as the measurement of a single parameter—related to the medium's properties—rather than a local reflectivity combining several parameters that is noninvertible. Fatemi and Kak[2] calculated this coefficient as a function of sound speed, though this was done by ignoring local density variation, which is an unreasonable assumption in medical imaging. We show that their coefficient can be interpreted as an approximation of a more general expression.

The expansion and solution of the wave equation proposed are very close to Jensen's approach;[8] consequently the results that have already been demonstrated in detail are used directly, and our contribution is highlighted. Jensen's development[8] was chosen because it offers flexible and widespread solutions for the simulation of ultrasonic imaging with various probe geometries, and, like the model of Zemp *et al.*,[7] it can also take into account both attenuation[11] and space-varying PSFs.

The paper is organized as follows: First, the general wave equation is modified to establish the expression to be solved under the assumptions chosen. An expression of the received signal is established in Sec. III and the results are discussed in Sec. IV.

II. DERIVATION OF THE WAVE EQUATION

Assuming that the particle velocity in the medium is small, which is usually the case at the power dissipated by medical ultrasound scanners, the wave equation can be written as[8]

$$\nabla^2 p_1(\boldsymbol{r},t) - \frac{1}{c(\boldsymbol{r})^2}\cdot\frac{\partial^2 p_1(\boldsymbol{r},t)}{\partial t^2} = \frac{1}{\rho(\boldsymbol{r})}\,\boldsymbol{\nabla}\,\rho(\boldsymbol{r})\cdot\boldsymbol{\nabla} p_1(\boldsymbol{r},t), \tag{1}$$

where p_1 is the pressure perturbation due to the propagation, at location $\boldsymbol{r}$ (Fig. 1) and time t, ρ is the local density, and c is the local speed of sound. Bold characters stand for three-dimensional triplets. $\boldsymbol{\nabla}$ is the gradient operator and ∇^2 is the Laplace operator. The solution method consists in isolating all the terms expressing a propagating wave at a constant speed on the left-hand side of the equation, and keeping the others on the right-hand side, which stands for the scattering operator. Since it assumes that the particle velocity is small in the insonified medium, this development is only suitable for linear imaging and does not take into account harmonic tissue imaging[12,13] and contrast agent imaging.[14–17] In the later developments, when no ambiguity is possible, the function's variables will be omitted for purposes of legibility.

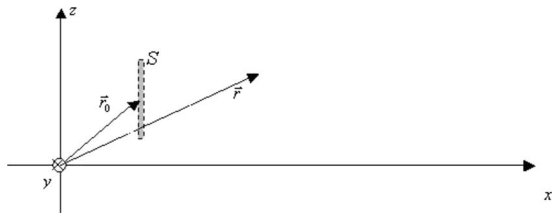


FIG. 1. Definition of the transducer's surface (S), position (r_0), and the transmission/scattering location $\boldsymbol{r}$.

Reorganizing the left- and right-hand terms of Eq. (1) gives

$$\underset{c_0}{\Xi} p_1 = Fop[p_1] \tag{2}$$

with

$$\underset{c_0}{\Xi} = \nabla^2(\cdot) - \frac{1}{c_0^2}\cdot\frac{\partial^2}{\partial t^2}(\cdot) \tag{3}$$

the D'Alembert operator in c_0 and

$$Fop = \frac{c^2}{c_0^2}\cdot\frac{\boldsymbol{\nabla}\rho}{\rho}\cdot\boldsymbol{\nabla}(\cdot) - \left(\frac{c^2}{c_0^2} - 1\right)\cdot \mathrm{div}(\boldsymbol{\nabla}(\cdot)) \tag{4}$$

the scattering operator. The ultrasonic process is described as the propagation of a progressive wave [left member of Eq. (2)] at constant speed c_0, which is perturbed by local inhomogeneities (described in the right member). However, contrary to Ref. 8, no first-order expression is made of $c(\boldsymbol{r})$ and $\rho(\boldsymbol{r})$ for the elaboration of the scattering operator Fop, given in Eq. (4), even if it will be developed considering small perturbations. Equation (2) is the equation that has to be solved in order to calculate the expression of the received signal.

III. CALCULATION OF THE RECEIVED SIGNAL

The solution is obtained by assuming that the incident field is a solution of the homogeneous part of Eq. (2) and its expression is established following Jensen's method.[8] Then the scattered field is calculated under the first-order Born approximation, and its integration over the transducer surface finally gives the expression of the received signal, which is a function of the local scattering coefficient.

A. Calculation of the incident field

The incident field propagating in tissues is assumed to be an eigensolution of the homogeneous part of Eq. (2) and is assumed to derive from a velocity potential ψ in Eq. (5). It is implicitly expressed by Eq. (6) as follows:

$$\boldsymbol{v}(\boldsymbol{r},t) = \nabla\psi(\boldsymbol{r},t), \tag{5}$$

$$\nabla p_i(\boldsymbol{r},t) = -\rho \cdot \frac{\partial \boldsymbol{v}(\boldsymbol{r},t)}{\partial t}. \tag{6}$$

Combining Eqs. (5) and (6) gives

$$\nabla p_i = -\rho \cdot \frac{\partial \nabla \psi}{\partial t}, \tag{7}$$

where $\psi(\boldsymbol{r},\boldsymbol{r}_0,t)$ is given at $\boldsymbol{r}$ for a single transducer in $\boldsymbol{r}_0$ by the following expression:[2]

$$\psi(\boldsymbol{r},\boldsymbol{r}_0,t) = v_n(t) \underset{t}{*} h(\boldsymbol{r},\boldsymbol{r}_0,t), \tag{8}$$

where $\underset{t}{*}$ denotes time convolution, and

$$h(\boldsymbol{r},\boldsymbol{r}_0,t) = -\int_{\boldsymbol{q}\in S} \frac{\delta\left(t - \frac{|\boldsymbol{r}-\boldsymbol{r}_0+\boldsymbol{q}|}{c_0}\right)}{4\cdot\pi\cdot|\boldsymbol{r}-\boldsymbol{r}_0+\boldsymbol{q}|} \cdot d^2\boldsymbol{q}, \tag{9}$$

where $v_n(t)$ is the normal speed of the transducer's surface S, and $h(\boldsymbol{r},\boldsymbol{r}_0,t)$ is the PSF in transmission for a transducer element placed in $\boldsymbol{r}_0$.

Then, Eq. (7) gives

$$\nabla p_i(\boldsymbol{r},\boldsymbol{r}_0,t) = -\rho \cdot \frac{\partial \nabla \psi(\boldsymbol{r},\boldsymbol{r}_0,t)}{\partial t}$$

$$= -\rho \cdot \frac{\partial v_n(t)}{\partial t} \underset{t}{*} \nabla h(\boldsymbol{r},\boldsymbol{r}_0,t). \tag{10}$$

The hypothesis of constant sound speed has been made for propagation (i.e., $c=c_0$), and we have assumed that $v_n(t)$ is constant over the transducer's surface. Equation (10) does not directly provide an expression of the incident field p_i, but this implicit expression is sufficient for the expression of the received signal. The use of ρ in this equation, and not ρ_0 as in Ref. 8, is discussed in Sec. IV.

B. Calculation of the scattered field

The pressure at $\boldsymbol{r}$ is calculated by integrating the ambient field p_1 over the volume space V. This is done using Green's function[18] and leads to the scattered field[8]

$$p_s(\boldsymbol{r},\boldsymbol{r}_0,t) = \int_V \int_T Fop[p_1(\boldsymbol{\xi},\boldsymbol{r}_0,\tau)] \cdot G(\boldsymbol{\xi},\tau|\boldsymbol{r},t) \cdot d\tau \cdot d^3\boldsymbol{\xi}, \tag{11}$$

where integration is through space ($\boldsymbol{\xi}\in V$) and time ($\tau\in T$), and G is Green's function of the homogeneous equation [Eq. (2)]

$$G(\boldsymbol{\xi},\tau|\boldsymbol{r},t) = -\frac{\delta\left(t-\tau-\frac{|\boldsymbol{r}-\boldsymbol{\xi}|}{c_0}\right)}{4\cdot\pi\cdot|\boldsymbol{r}-\boldsymbol{\xi}|}. \tag{12}$$

Under the first-order Born approximation, it is shown[8] that p_1 is close to the incident field p_i, and p_s becomes

$$p_s(\boldsymbol{r},\boldsymbol{r}_0,t) = \int_V \int_T Fop[p_i(\boldsymbol{\xi},\boldsymbol{r}_0,\tau)] \cdot G(\boldsymbol{\xi},\tau|\boldsymbol{r},t) \cdot d\tau \cdot d^3\boldsymbol{\xi}. \tag{13}$$

p_s now has to be integrated over the transducer's surface in order to express the received signal.

C. Calculation of the received signal

The received signal is the scattered pressure field p_s integrated over the transducer's surface S, convolved with the electromechanical impulse response $E_m(t)$ of the transducer,

$$p_r(\boldsymbol{r}_0',\boldsymbol{r}_0,t) = E_m(t) \underset{t}{*} \int_S p_s(\boldsymbol{r}_0'+\boldsymbol{\xi},\boldsymbol{r}_0,t) d^2\boldsymbol{\xi} \tag{14}$$

for a receiving transducer placed at point r_0'. Then, considering that

$$p_s(\boldsymbol{r},\boldsymbol{r}_0,t) = \int_V \int_T Fop[p_i(\boldsymbol{\xi},\boldsymbol{r}_0,\tau)] \cdot G(\boldsymbol{\xi},\tau|\boldsymbol{r},t) \cdot d\tau \cdot d^3\boldsymbol{\xi} \tag{15}$$

$$= \int_V \int_T Fop_2[\nabla p_i(\boldsymbol{\xi},\boldsymbol{r}_0,\tau)] \cdot G(\boldsymbol{\xi},\tau|\boldsymbol{r},t) \cdot d\tau \cdot d^3\boldsymbol{\xi} \tag{16}$$

with

$$Fop_2 = \frac{c^2}{c_0^2} \cdot \frac{\nabla\rho}{\rho} \cdot (\cdot) - \left(\frac{c^2}{c_0^2} - 1\right) \cdot \mathrm{div}(\cdot), \tag{17}$$

inserting Eqs. (10) and (16) into Eq. (14) leads to

$$p_r(\boldsymbol{r},\boldsymbol{r}_0,t) = -\frac{E_m(t)}{4} \underset{t}{*} \frac{\partial v(t)}{\partial t} \underset{t}{*} \int_V Fop_2[\rho \cdot \nabla h(\boldsymbol{\xi},\boldsymbol{r}_0,t)]$$

$$\underset{t}{*} h'(\boldsymbol{r}_0,\boldsymbol{\xi},t) \cdot d^3\boldsymbol{\xi}, \tag{18}$$

where $h'(r_0',\xi,t)$ is Green's function integrated over the transducer's surface, that is, the spatial impulse response of the transducer for reception. This expression is equivalent to

$$p_r(\boldsymbol{r},\boldsymbol{r}_0,t) = -\frac{E_m(t)}{4} \underset{t}{*} \frac{\partial v(t)}{\partial t} \underset{t}{*} \int_V Fop_3[h(\boldsymbol{\xi},\boldsymbol{r}_0,t)]$$

$$\underset{t}{*} h'(\boldsymbol{r}_0,\boldsymbol{\xi},t) \cdot d^3\boldsymbol{\xi} \tag{19}$$

with

$$Fop_3 = \frac{c^2}{c_0^2} \cdot \nabla\rho \cdot \nabla(\cdot) - \left(\frac{c^2}{c_0^2} - 1\right) \cdot \mathrm{div}(\rho \cdot \nabla(\cdot)). \tag{20}$$

This expression is approximated by assuming that the perturbation operator is, in fact, applied to the full PSF

$$pr(\mathbf{r},\mathbf{r}_0,t) = \frac{Em(t)}{4} *_t \frac{\partial v(t)}{\partial t} *_t \int_V Fop_3[h] *_t h' \cdot d^3\xi$$

$$\approx \frac{Em(t)}{4} *_t \frac{\partial v(t)}{\partial t} *_t \int_V Fop_3[h *_t h'] \cdot d^3\xi. \quad (21)$$

This approximation and its implications are discussed in Sec. IV.

Then, defining

$$h_{pe}(\mathbf{r},\mathbf{r}_0,t) = h(\mathbf{r},\mathbf{r}_0,t) *_t h'(\mathbf{r}_0,\mathbf{r},t) \quad (22)$$

can be written as

$$\int_V Fop_3[h_{pe}] \cdot d^3\xi = \int_V [\alpha \cdot \nabla\rho \cdot \nabla h_{pe} - (\alpha - 1) \cdot \mathrm{div}(\rho \cdot \nabla h_{pe})] \cdot d^3\xi, \quad (23)$$

where $\alpha(\mathbf{r}) = c(\mathbf{r})^2 / c_0^2$.

Then the Green–Ostrogradsky theorem and the properties of the divergence can be written as

$$\int_V \mathrm{div}(u \cdot \boldsymbol{f}) \cdot \mathrm{div} = \int_V (\nabla u \cdot \boldsymbol{f} + u \cdot \mathrm{div}(\boldsymbol{f})) \cdot \mathrm{div} = \int_{S_V} u \cdot \boldsymbol{f} \cdot \mathbf{n} \cdot ds \quad (24)$$

with $\mathbf{n}$ the outward normal vector to the surface S_V of the integration volume V. If $u \cdot \boldsymbol{f}$ decreases rapidly, the integral is null for an unbounded space, leading to the relationship

$$\int_V u \cdot \mathrm{div}(\boldsymbol{f}) \cdot \mathrm{div} = -\int_V \nabla(u) \cdot \boldsymbol{f} \cdot \mathrm{div}. \quad (25)$$

Assuming that h_{pe} varies slowly toward infinity and tends to zero, which is the case for ultrasound beams, and using Eq. (25), Eq. (23) becomes

$$\int_V Fop_3[h_{pe}] \cdot d^3\xi = \int_V [\alpha \cdot \nabla\rho \cdot \nabla h_{pe} + \nabla(\alpha - 1) \cdot \rho \cdot \nabla h_{pe})] \cdot d^3\xi, \quad (26)$$

$$= \int_V \nabla(\alpha \cdot \rho) \cdot \nabla h_{pe} \cdot d^3\boldsymbol{\xi}, \quad (27)$$

$$= -\int_V (\alpha \cdot \rho) \cdot \nabla^2 h_{pe} \cdot d^3\boldsymbol{\xi}, \quad (28)$$

and the final expression for the scattered field is

$$p_r(\mathbf{r}_0,t) = \rho_0 \cdot \frac{E_m(t)}{4} *_t \frac{\partial v(t)}{\partial t} *_t \int_V \left[\frac{c^2 \cdot \rho}{c_0^2 \cdot \rho_0}\right] \cdot \nabla^2 h_{pe}(\boldsymbol{\xi},\mathbf{r}_0,t) \cdot d^3\boldsymbol{\xi}, \quad (29)$$

$$\Leftrightarrow p_r(\mathbf{r}_0,t) = \rho_0 \cdot \frac{E_m(t)}{4} *_t \frac{\partial v(t)}{\partial t} *_t \int_V \nabla^2 \left[\frac{K(\boldsymbol{\xi})}{K_0}\right] \cdot h_{pe}(\boldsymbol{\xi},\mathbf{r}_0,t) \cdot d^3\boldsymbol{\xi}, \quad (30)$$

where $K = c^2 \cdot \rho$ is the longitudinal bulk modulus, which is the inverse of the compressibility modulus and in this very case equal to the longitudinal Young modulus,[3,4] and K_0 its average value such that

$$K(\mathbf{r}) = K_0 + \Delta K(\mathbf{r}), \quad (31)$$

assuming that h_{pe} is the solution of the homogeneous part of Eq. (2). Using Eq. (30) leads to

$$p_r(\mathbf{r}_0,t) = \frac{\rho_0}{2 \cdot c_0^2} \cdot E_m(t) *_t \frac{\partial^3 v(t)}{\partial t^3} *_t \int_V h_{pe}(\boldsymbol{\xi},\mathbf{r}_0,t) \cdot \left[\frac{K(\boldsymbol{\xi})}{K_0} + \varphi\right] \cdot d^3\boldsymbol{\xi}, \quad (32)$$

where $K/K_0 + \varphi$ is the scattering coefficient R. The term φ is a constant that is naturally discarded by the convolution form, since the imaging process is only sensitive to the variations in the local properties. It is introduced so that the scattering coefficient can be extracted from the general equation while retaining a physical meaning, but has no impact on the scattered pressure. As the scattering coefficient itself should be zero when $K = K_0$, the value for the constant is $\varphi = -1$, and the complete expression for R is

$$R = \frac{K - K_0}{K_0}. \quad (33)$$

Finally, the model expression can be reduced to

$$p_r(\mathbf{r}_0,t) = v_{pe}(t) *_t h_{pe}(\mathbf{r},\mathbf{r}_0,t) *_r R(\mathbf{r}), \quad (34)$$

where $*_r$ denotes spatial convolution and

$$v_{pe}(t) = \frac{\rho_0}{4c_0^2} \cdot E_m(t) *_t \frac{\partial^3 v(t)}{\partial t^3}, \quad (35)$$

$$h_{pe}(\mathbf{r},\mathbf{r}_0,t) = h(\mathbf{r},\mathbf{r}_0,t) *_t h(\mathbf{r}_0,\mathbf{r},t), \quad (36)$$

$$R(\mathbf{r}) = \frac{K(\mathbf{r}) - K_0}{K_0}, \quad (37)$$

$$K = c^2 \cdot \rho. \quad (38)$$

The term R in Eq. (37) is the scattering term, which is defined here as the relative radial bulk modulus. This expression is valid only for small variations in the bulk modulus around its average value, that is, with $\Delta K(\mathbf{r}) \ll K_0$, if $\Delta K(\mathbf{r})$ is the variation of K at point $\mathbf{r}$. As in the previous model, the speed of sound and density variations cannot be estimated independently; however, under the assumptions, in the scattering coefficient, they form the expression of another local parameter: longitudinal bulk modulus.

IV. DISCUSSION

The proposed model is developed under the same main hypotheses as the usual linear descriptions of the ultrasonic imaging process,[1,2,7,8] essentially linear propagation and the Born approximation. The main difference is the use of non-approximated forms for c and ρ, but other assumptions have been modified or adapted.

A. Expression of the incident field

The expression of the incident field used in this model is slightly different from Jensen's expression.[8] We hypothesize that it is possible to consider that the density ρ is fully unknown at the measurement point and thus not "partially" equal to ρ_0, as implied by Eq. (27) in Ref. 8. Indeed, in Ref. 8, ρ_0 is the density at $\boldsymbol{r}$ where density is assumed to be unknown. The present hypothesis is not newly introduced because once the influence of sound speed and density on the propagation are considered separately, and once a linear behavior of the system is assumed (which means that the response of the scattering coefficient collection is the sum of the responses of each one), there is no need to consider, at a certain point $\boldsymbol{r}$, the past density along the propagation profile, which is implicitly set to ρ_0. Here the field is assumed to be known until the ultrasound pulse reaches inhomogeneity, where the density and sound speed variations are then fully taken into account. This appears in the model being developed through the use of ρ in Eq. (6). The solution was performed assuming that the propagation medium's density and speed of sound are equal to their average values ρ_0 and c_0 everywhere, scattering point $\boldsymbol{r}$ included, but also that these quantities are varying at this point. This contradiction regarding the density changes is moved here by making the same hypothesis until the incidence on a measurement point $\boldsymbol{r}$, but *not* at this point. The difference is that the pressure changes at point $\boldsymbol{r}$ are more accurately estimated, through a better use of the assumptions required to assume that the incoming pressure is known.

On the other hand, the use of c_0 in Eqs. (9) and (12) emphasizes that a linear development includes contradictions, because this assumption of propagation at the mean speed of sound includes the estimation point $\boldsymbol{r}$. This last problem cannot be avoided here. A more complete description should imply the local variations in the speed of sound so as to take into account phase changes during propagation, as done very recently in Ref. 19.

B. Approximation of Equation (21)

The development proposed is possible by considering that the scattering operator is taken as Eq. (21) on the entire PSF, including the time convolution with the reception profile h'. This operation expresses the typical assumption of linear approaches that the received signal is the result of a convolution with h'. Indeed, the linearization of the initial equations implies that the resulting model will express the received pressure as a function of a convolution product with the PSF. It could also be expressed by taking h' out of the volume integral, ignoring the spatial aspect of the convolution (for h'), and considering that h' is applied along the propagation axis (or ultrasonic line). This approximation is equivalent to saying that h' can be roughly viewed solely as a function of time. In this case, the time convolution makes it possible to move it at will inside the equation. It is also equivalent to assuming that the local variations of the medium properties are concomitant, so that the terms that are only a function of density can be expressed by its mean value. Following this point of view, consulting a table giving the speed of sound and density of different biological materials[6] shows that the variations in the speed of sound in most cases imply a variation in density.

This assumption would imply in the volume integration that a term that is a function of the convolution of the gradients of the transmission and reception functions h and h' is removed, which underlines the approximate aspect of such approaches and the importance of slow variations of h and h' over the integration volume. It should also be noted that this is equivalent to Jensen's approximation,[8] when he assumed that h_{pe} is the solution of the homogeneous part of the wave equation, which consists in considering that h can be replaced with h_{pe} straightforwardly, discarding the additional terms that would appear when convolving the D'Alembert of h by h'.

C. Expression of the established scattering coefficient

Equations (37) and (38) clearly show the influence of the speed of sound c on the scattering coefficient. A square dependence is obtained, in coherence with Fatemi and Kak's[2] results. The expression of R as a function of bulk modulus K can also be linked to Gore and Leeman's[1] result, and to the result of Zemp *et al.*,[7] which are partially a function of the local compressibility. Replacing c and ρ in Eq. (37) with Jensen's development [Ref. 8, Eq. (12)] expressed in Eq. (39) we have

$$c = c_0 + \Delta c,$$

$$\rho = \rho_0 + \Delta\rho \tag{39}$$

with $\Delta c \ll c_0$ and $\Delta\rho \ll \rho_0$, and ignoring the higher-order terms gives

$$R \approx \frac{\Delta\rho}{\rho_0} + \frac{2\Delta c}{c_0}, \tag{40}$$

which is, except for the minus sign, Jensen's expression [Ref. 8, Eq. 47]

$$R_{\text{Jensen}} = \frac{\Delta\rho}{\rho_0} - \frac{2\Delta c}{c_0}. \tag{41}$$

The negative dependence of Jensen's scattering coefficient[8] on the speed of sound variations Δc does not seem to reflect a physical property of reflection because it implies that a positive variation of c would reduce the local reflectivity or generate a negative scattering coefficient. Indeed, the nonlinear scattering coefficient for normal incidence $R_{\text{NL}} = (Z - Z_0)/(Z + Z_0)$ shows that an increase in the local mechanical properties results in a positive scattering coefficient. Checking both developments suggests that a sign change is missing

between Jensen's Eqs. (40) and (41) in Ref. 8. Fortunately, this sign mistake has no real impact on the extensive use of the simulator[9] based on these equations, since Eq. (41) is not usually applied, the scattering coefficient R_{Jensen} being directly defined instead. Finally, the comparison of Eqs. (40) and (41) suggests that R_{Jensen} can be seen as a partial development of the expression of the scattering term R, and should be modified such that

$$R^{+}_{\text{Jensen}} = \frac{\Delta\rho}{\rho_0} + \frac{2\Delta c}{c_0} \tag{42}$$

is the corrected Jensen coefficient.[8]

It should be noticed that transforming the corrected Jensen coefficient[8] would lead back to

$$R^{+}_{\text{Jensen}} = \frac{\Delta\rho}{\rho_0} + \frac{2\Delta c}{c_0} = \frac{c_0^2 \cdot \Delta\rho + 2 \cdot \rho_0 \cdot c_0 \cdot \Delta c}{\rho_0 \cdot c_0^2} \tag{43}$$

with the expression of $\rho_0 \cdot c_0^2 = K_0$ at the denominator underlining the closeness with the proposed model, while the numerator is like the first-order expression of K.

The expression of the scattering coefficient as function of the bulk modulus K may be counterintuitive. On another hand, some assumptions such as the fact that the medium is considered as homogeneous until scattering occurs is also counterintuitive when considering imaging biological tissues, as well as the assumption that the waves are traveling at constant speed. The proposed model merely takes the typical set of assumptions close to their limits, showing their real impact on the expression of the scattering coefficient.

Figure 2 displays the values of the proposed scattering coefficient R and the corrected Jensen coefficient R^{+}_{Jensen}.[8] Figure 2(a) illustrates the behavior of the scattering coefficients when the density is set at ρ_0 and Fig. 2(b) illustrates the behavior when the speed of sound is set at c_0. In this latter case, both coefficients are equal, but when density is set at ρ_0 [Fig. 2(a)], the R and R^{+}_{Jensen} curves are tangential in c_0, showing their equivalence for small variations. However, R remains greater than -1 when c tends toward zero. The model presented in this paper is not quantitatively valid for large local variations K, but one can see in Eq. (37) that R also tends to -1 as K tends to zero and retains a physical meaning ($R > -1$), which is advantageous for a linear approach of such cases. Unfortunately, the model suffers from its approximations for strong positive variations of K above K_0, and Eq. (37) does not have any physical signification for $K > 2 \cdot K_0$, slightly sooner than R^{+}_{Jensen}, with the scattering coefficient then greater than 1. Nevertheless, the result given here provides an interesting approach for linearly modeling large local decreases in the mechanical properties of the tissue below their mean values. Those large decreases could reflect the existence of local very light inhomogeneities in the propagating medium, such as large bubbles or local vacuums.

For the simulation of such large variations, if needed, one should use the following definition to compute the scattering coefficient or to interpret the estimation results of local reflectivity:

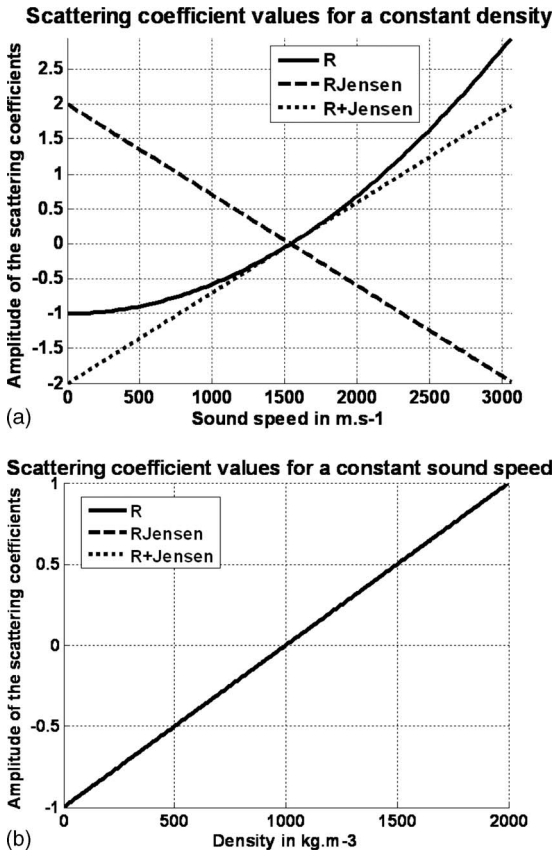


FIG. 2. Illustration of the dependence of the relative bulk modulus scattering coefficient R on changes in speed of sound (a) and density (b) separately, compared to the values provided by the Jensen coefficient R_{Jensen} (Ref. 8) and its modified version R^{+}_{Jensen}.

$$R = \frac{K - K_0}{K_0}, \quad K \in [0;2K_0],$$

$$R = 1, \quad K > 2K_0. \tag{44}$$

Equation (44) underlines that, in the model proposed, when K is greater than $2 \cdot K_0$, the solution of the inverse linear approach cannot distinguish dense matter and systematically underestimates the values of their mechanical properties. However, below K_0, the proposed model is coherent and more reliable [Fig. 2(a)] than Jensen's coefficient.[8]

Given that these two models—the one proposed by Jensen[8] and the one proposed in this paper—have very close, nearly equal expressions and given that their scattering coefficients, once corrected, are tangential around their mean value, they have the same ability to simulate the ultrasonic imaging process.

V. CONCLUSION

An expansion of the wave equation for propagation in an inhomogeneous medium has been established under the assumption of linear propagation. Solving it under the classical assumptions leads to an expression of the received rf signal that describes the local scattering coefficient as the variation of a single mechanical parameter. This model shows that considering the scattering as a function of small variations in density and speed of sound around their mean values is

equivalent to associating the acoustical acquisition with a measure of the relative longitudinal bulk modulus. The model developed thus offers a way to estimate the bulk modulus of the imaged matter through the solution of the inverse problem in the case of small local variations. The final expression of the scattering coefficient is shown to be coherent with previous developments, while the model proposed is very close to the one presented by Jensen in 1991 (Ref. 8) and now widely used for ultrasonic imaging simulation. The similarity of the two models also implies that our model presents all the abilities of Jensen's model[8] for simulating ultrasonic imaging and is more accurate when the scattering coefficient distribution is derived from the speed of sound and density fields. Moreover, the expression proposed is particularly convenient for the linear simulation of extremely soft inclusion imaging. The corresponding scattering coefficient, even if a linear model is intrinsically not well suited for such uses, retains a physical meaning when density or speed of sound decreases, even to zero.

[1]J. C. Gore and S. Leeman, "Ultrasonic backscattering from human tissue: A realistic model," Phys. Med. Biol. **22**, 317–326 (1977).
[2]M. Fatemi and A. C. Kak, "Ultrasonic B-scan imaging: Theory of image formation and a technique for restoration," Ultrason. Imaging **2**, 1–47 (1980).
[3]T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994), p. 100.
[4]R. T. Beyer and S. V. Letcher, *Physical Ultrasonics* (Academic, New York, 1969).
[5]J. M. Mari, "Localisation échographique d'inclusions fortement échogènes en tissus mous, application à la détection d'électrode (Ultrasonic localization of strongly echogeneous inclusions in soft tissues, application to micro needle detection)," Ph.D. thesis, Université Claude Bernard Lyon 1, France (2004).
[6]D. Royer and E. Dieulesaint, *Ondes Élastiques Dans les Solides (Elastic Waves in Solids)* (Masson, Paris, 1996), Vol. **1**.
[7]R. Zemp, C. K. Abbey, and M. F. Insana, "Linear system models for ultrasonic imaging: Application to signal statistics," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 642–654 (2003).
[8]J. A. Jensen, "A model for the propagation and scattering of ultrasound in tissue," J. Acoust. Soc. Am. **89**, 182–190 (1991).
[9]J. A. Jensen, "Field: A program for simulating ultrasound systems," Med. Biol. Eng. Comput. **34**, 351–354 (1996).
[10]J. A. Jensen, "Deconvolution of ultrasound images," Ultrason. Imaging, **14**, 1–15 (1992).
[11]J. A. Jensen, D. Gandhi, and W. D. O'Brien, "Ultrasound fields in an attenuating medium," Proceedings of the IEEE 1993 Ultrasonics Symposium (1993), Vol. **2**, pp. 943–946.
[12]M. E. Frijlink, D. E. Goertz, A. Bouakaz, and A. F. W. van der Steen, "Intravascular ultrasound tissue harmonic imaging: A simulation study," Ultrasonics, **44**, e185–e188 (2006).
[13]F. Tranquart, N. Grenier, V. Eder, and L. Pourcelot, "Clinical use of ultrasound tissue harmonic imaging," Ultrasound Med. Biol. **25**, 889–894 (1999).
[14]P. J. A. Frinking, A. Bouakaz, J. Kirkhorn, F. J. Ten Cate, and N. de Jong, "Ultrasound contrast imaging: Current and new potential methods," Ultrasound Med. Biol. **26**, 965–975 (2000).
[15]M. Tang and R. Eckersley, "Frequency dependent attenuation and scattering by microbubbles," Ultrasound Med. Biol. **33**, 164–168 (2007).
[16]M. Tang, R. Eckersley, and J. Noble, "Pressure-dependent attenuation with microbubbles at low mechanical index," Ultrasound Med. Biol. **31**, 377–384 (2005).
[17]J. M. Mari, K. Hibbs, and M. X. Tang, *A Nonlinear Ultrasonic Scattering Approach for Micro Bubble Concentration Quantification* (IEEE EMBC, Lyon, 2007).
[18]D. G. Duffy, *Green's Function With Applications* (Chapman and Hall/CRC, Boca Raton, FL, 2001), pp. 99–101.
[19]D. Napolitano, C.-H. Chou, G. McLaughlin, T.-L. Ji, L. Mo, D. DeBusschere, and R. Steins, "Sound speed correction in ultrasound imaging," Ultrasonics **44**, e43–e46 (2006).

Magnetic resonance imaging of boiling induced by high intensity focused ultrasound

Tatiana D. Khokhlova
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, 1013 NE 40th, Seattle, Washington 98105 and International Laser Center, Moscow State University, Moscow, 119992, Russian Federation

Michael S. Canney
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, 1013 NE 40th, Seattle, Washington 98105

Donghoon Lee and Kenneth I. Marro
Department of Radiology, School of Medicine, University of Washington, Seattle, Washington 98105

Lawrence A. Crum
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, 1013 NE 40th, Seattle, Washington 98105

Vera A. Khokhlova
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, 1013 NE 40th, Seattle, Washington 98105 and Department of Acoustics, Physics Faculty, Moscow State University, Moscow, 119992, Russian Federation

Michael R. Bailey[a)]
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, 1013 NE 40th, Seattle, Washington 98105

(Received 20 August 2008; revised 12 January 2009; accepted 20 January 2009)

Both mechanically induced acoustic cavitation and thermally induced boiling can occur during high intensity focused ultrasound (HIFU) medical therapy. The goal was to monitor the temperature as boiling was approached using magnetic resonance imaging (MRI). Tissue phantoms were heated for 20 s in a 4.7-T magnet using a 2-MHz HIFU source with an aperture and radius of curvature of 44 mm. The peak focal pressure was 27.5 MPa with corresponding beam width of 0.5 mm. The temperature measured in a single MRI voxel by water proton resonance frequency shift attained a maximum value of only 73 °C after 7 s of continuous HIFU exposure when boiling started. Boiling was detected by visual observation, by appearance on the MR images, and by a marked change in the HIFU source power. Nonlinear modeling of the acoustic field combined with a heat transfer equation predicted 100 °C after 7 s of exposure. Averaging of the calculated temperature field over the volume of the MRI voxel ($0.3\times0.5\times2$ mm^3) yielded a maximum of 73 °C that agreed with the MR thermometry measurement. These results have implications for the use of MRI-determined temperature values to guide treatments with clinical HIFU systems.
 [DOI: 10.1121/1.3081393]

PACS number(s): 43.80.Gx [CCC] Pages: 2420–2431

I. INTRODUCTION

The use of high intensity focused ultrasound (HIFU) for tissue ablation is rapidly achieving clinical acceptance for a wide range of medical applications.[1–4] The tissue ablation mechanism in the majority of these applications is coagulative necrosis, induced by heating of the tissue due to absorption of the intense ultrasound. Heating is often monitored by magnetic resonance imaging (MRI), in which the tissue temperature is calculated from direct measurements of the MR-signal phase change resulting from the water proton resonance frequency shift.[5] The temperature measurements are then used to calculate the region of necrosed tissue—the lesion—based on a thermal dose criterion.[6] MRI monitoring and lesion determination are used in the only clinical, transcutaneous HIFU device approved by the United States Food and Drug Administration (FDA).[7] New treatment protocols are also being developed on this and similar machines that use high-amplitude ultrasonic pulses on the presumption of creating cavitation bubbles in tissue for enhanced heating.[8–11] However, enhanced heating can also occur due to nonlinear propagation effects, and such heating may be sufficient to cause boiling bubbles.[12] When bubbles and enhanced heating have been observed *in vivo*, the MRI-determined temperature was less than 100 °C, and the result

a)Author to whom correspondence should be addressed. Electronic mail: bailey@apl.washington.edu

 0001-4966/2009/125(4)/2420/12/$25.00

was interpreted as cavitation-enhanced heating.[10] We speculate that the peak temperature may have been higher and that bubbles were due to boiling, the consequence, not the cause, of enhanced heating.

Acoustic cavitation and nonlinear acoustic propagation are two nonlinear mechanisms that can enhance HIFU heating beyond that expected to be produced by absorption at the HIFU excitation frequency. Nonlinear acoustic propagation distorts the HIFU wave and causes acoustic energy to be pumped from the fundamental excitation frequency to higher harmonic frequencies. Higher frequencies are more readily absorbed than lower ones and thus generate enhanced heating.[12–14] An additional effect of nonlinear propagation is that the extra heating and initial boiling, if attained, are more localized than would be expected assuming linear propagation conditions.[14,15] One way to increase nonlinear effects is to increase the acoustic pressure as this accelerates the waveform distortion. With a sufficient pressure amplitude, nonlinear acoustic propagation results in a shock wave, which contains hundreds of harmonics and can cause boiling in milliseconds.[16]

Cavitation bubbles are nonlinear scatterers that result in acoustic re-radiation at frequencies higher than the fundamental HIFU wave and therefore generate enhanced HIFU heating.[9,11] The bubbles may also cause heating by viscous damping of their oscillations[17] and by diffusion of heat from their gaseous interiors that are heated when compressed in oscillation.[18] Increasing acoustic pressure amplitude may increase cavitation-enhanced heating as more bubble nuclei are recruited and bubbles are driven into more violent oscillation. Cavitation-enhanced heating has been demonstrated in tissue-mimicking phantoms, where the cavitation threshold is likely lower than in tissue, and only at low HIFU pressures (peak negative pressure amplitude <4 MPa).[9,19] Within this range, a small increase in HIFU pressure amplitude led to both detectable cavitation activity and significantly greater heating measured by thermocouples.[9,19] Other observations of enhanced heating that were attributed to cavitation are summarized in Ref. 9.

Cavitation and boiling bubbles created by HIFU are visibly distinct when observed in optically transparent gel, tissue phantoms.[16] The cavitation or mechanically generated bubbles are diffuse and micron-sized, whereas the boiling or thermally generated bubbles are focal and millimeter-sized. Given sufficient acoustic pressures within the clinical range, both types of bubbles can appear quickly—cavitation in microseconds and boiling in milliseconds.[16,20] Several acoustic techniques have been reported to differentiate the two phenomena.[15,16,21,22] However, few of these techniques have yet been applied in bio-effect studies.

In bio-effect studies, in particular, all bubble activity is often categorized as cavitation, and the appearance of a bubble is a treatment-altering event.[23] Bubbles cause backscatter of HIFU which results in distortion and migration of the lesion,[24] scatter of imaging ultrasound that can be detected and used for guidance,[25] and mechanical erosion of tissue.[26] *In vitro* studies suggest that the contribution of cavitation in at least the first two of these effects—acoustic backscatter and lesion distortion—is negligible compared to that of boiling, which might be expected since boiling bubbles are much larger.[15] The tissue ablation arising from boiling and cavitation bubbles may also be different because of their size, motion, and surrounding temperature. Whether the clinical goal is to use cavitation or boiling bubbles or to avoid them in treatments, it is necessary to understand how the bubbles are created.

One example is to determine how enhanced heating and the presence of bubbles are related. A study that is commonly cited in the HIFU literature of cavitation-enhanced heating *in vivo* is that by Sokka *et al.*[10] In this study, a short (0.5-s) high-amplitude pulse, preceding a long low-amplitude pulse, caused a greater MR-measured temperature rise compared to only a long low-amplitude pulse. The authors may have assumed that the detected bubble activity was not boiling (and therefore bubbles were the *cause* not the *result* of enhanced heating) because the temperature measured by MRI did not reach 100 °C. This threshold temperature for HIFU-induced boiling to occur in tissue has been justified by calculations[16,22] and thermocouple measurements.[22,27] Although bubbles have been detected in MRI measurements in tissue, there have been no results reporting the use of MRI to detect HIFU-induced boiling and to distinguish boiling from cavitation.

The goal of this research was to use MR as an imaging technique to observe boiling during HIFU exposure and simultaneously to use MR thermometry to measure the temperature when boiling occurred. The accuracy of MRI temperature measurements was investigated by comparing the time to boil and the temperature rise measured by MR to those determined using other experimental techniques as well as theoretical modeling based on the Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation. The study was designed to enable a high MRI resolution and to work under well-controlled experimental conditions. Prior to the MR experiments, the acoustic field of the HIFU source was characterized, the cavitation pressure threshold and the temperature when boiling started in the phantoms were determined, and numerical simulations of the acoustic and temperature fields were performed.

II. THEORY

The temperature rise in the gel tissue phantom was numerically modeled by coupling an acoustic propagation model with a heat transfer model. The model equations and solution techniques have been described in previous publications and are only briefly summarized here.[12,15] A comparison of simulations and measurements of the acoustic field in water and gel for a source nearly identical to the one used here has also been reported.[28]

The HIFU field was modeled using a KZK-type nonlinear parabolic equation,[15] generalized for the frequency-dependent absorption properties of the propagation medium:

$$\frac{\partial}{\partial \tau}\left(\frac{\partial p}{\partial z}-\frac{\beta}{\rho_0 c_0^3}p\frac{\partial p}{\partial \tau}-L_{\mathrm{abs}}(p)\right)=\frac{c_0}{2}\Delta_{\perp}p. \quad (1)$$

Here p is the acoustic pressure, z is the propagation coordinate along the axis of the beam, $\tau=t-z/c_0$ is the retarded

time, c_0 is the sound speed, ρ_0 is the ambient density of the medium, β is the coefficient of nonlinearity, $\Delta_\perp = \partial^2/\partial r^2 + r^{-1}\partial/\partial r$ is the Laplacian with respect to the transverse coordinate r, and L_{abs} is the linear operator that accounts for absorption and dispersion in the medium.

The propagation of ultrasound was through a two-layer medium, consisting of water and gel. For water, the thermoviscous absorption was included as

$$L_{\text{abs}} = \frac{b}{2c_0^3\rho_0}\frac{\partial^2 p}{\partial \tau^2}, \tag{2}$$

where b is the dissipative parameter of water. For the tissue phantom, the operator L_{abs} accounted for the measured power law of absorption:

$$\alpha(f) = \alpha_0(f/f_0)^\eta, \tag{3}$$

where α_0 is the absorption coefficient at the fundamental frequency f_0. Small variation in the sound speed with frequency was calculated from the absorption law, Eq. (3), using local dispersion relations.[12]

The boundary condition for Eq. (1) was determined using the combined modeling and measurement calibration technique developed in a previous paper.[28] Equation (1) was then solved numerically in the frequency-domain. First, the acoustic pressure waveform was represented as a Fourier series expansion. Next, the set of nonlinear, coupled differential equations for the amplitudes of the harmonics were derived and integrated numerically using the method of fractional steps with an operator-splitting procedure.[15] Acoustic waveforms $p(\tau,z,r)$, spatial distributions of the intensities I_n of the harmonics nf_0, and total intensity of the wave

$$I(z,r) = \sum_{n=1}^{\infty} I_n(z,r) \tag{4}$$

were calculated. The distribution of heat sources q_v due to absorption of ultrasound,

$$q_v(z,r) = 2\sum_{n=1}^{\infty} \alpha(nf_0)I_n(z,r), \tag{5}$$

was obtained for further simulation of the temperature rise in the phantom.

The values of the physical constants used for acoustic modeling were $\rho_0 = 1000$ kg/m^3, $c_0 = 1486$ m/s, $\beta = 3.5$, and $b = 4.33 \times 10^{-3}$ kg s^{-1} m^{-1} for water and $\rho_0 = 1044$ kg/m^3, $c_0 = 1544$ m/s, $\beta = 4.0$, and $\alpha_0 = 1.6$ m^{-1} at 1 MHz, $\eta = 1$, for the tissue phantom.[29] Changes in the acoustic parameters of the phantom due to HIFU-induced heating were not considered in the simulations.

To quantify the effect of acoustic nonlinearity under the experimental conditions employed in this study, simulations were also performed assuming linear HIFU propagation by setting $\beta = 0$ in Eq. (1). In linear simulations, the HIFU waveforms remained sinusoidal; the intensity I_L included the intensity of the first harmonic only,

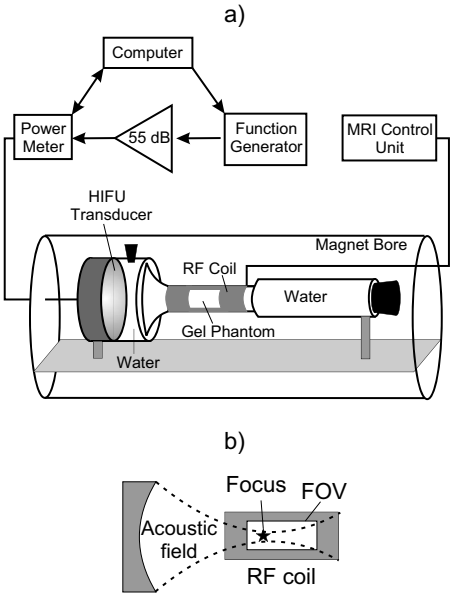


FIG. 1. Experimental arrangement (a) and relative position of the transducer focus and the MRI field of view (FOV) (b).

$$I_L(z,r) = I_1(z,r), \tag{6}$$

and corresponded to the more common but less accurate linearly derated intensity typically used in HIFU studies. The distribution of heat sources was calculated as twice the product of the intensity and the absorption coefficient at the source operating frequency

$$q_v(z,r) = 2\alpha(f_0)I_L(z,r). \tag{7}$$

The temperature rise in the phantom was modeled using the heat transfer equation

$$\frac{\partial T}{\partial t} = k\Delta T + \frac{q_v}{c_v}, \tag{8}$$

where T is the temperature in the phantom, c_v is the heat capacity per unit volume, k is the thermal diffusivity, and q_v is the distribution of thermal sources obtained from either nonlinear, Eq. (5), or linear, Eq. (7), acoustic simulations. Equation (8) was integrated numerically using finite differences.[12] The thermal properties of the phantom used in simulations were $c_v = 5.3 \times 10^6$ J m^{-3} °C^{-1} and $k = 1.3 \times 10^{-7}$ m^2/s.[29]

III. EXPERIMENTAL METHODS

The experimental arrangement is shown in Fig. 1. The transducer, coupling medium (degassed water), and tissue phantom were housed in a custom-designed cylindrical acrylic enclosure that was centered within the bore of the magnet. The 5-cm-long, 2.5-cm-diameter phantom sample was positioned at the focus of the transducer in the narrowest part of the enclosure, which was wrapped by an Alderman–Grant type radiofrequency (rf) volume coil. The phantom was narrow and could therefore be placed in a small volume coil, which provided a good filling factor, increased the signal-to-noise ratio, and optimized the spatial and temporal resolution. A water-filled Tygon tube capped with an acous-

tically absorptive rubber stopper was placed distal to the coil to prevent reflections in the experimental arrangement. The driving electronics were located outside the magnet room and consisted of a laptop computer running LABVIEW (National Instruments, Austin, TX), an HP33150 function generator (Palo Alto, CA), and an ENI A150 amplifier (Bloomington, NY).

The HIFU transducer had an aperture and radius of curvature of 44 mm and a resonant frequency of 2.185 MHz, and was mounted within the wall of the acrylic enclosure. The source was identical to one previously characterized,[28] but the housing was made of polycarbonate instead of metal to be MRI-compatible. The experimental exposure was continuous for 20 s. The electrical power delivered to the source was 63 W, and the acoustic power, measured by radiation force balance, was 49 W.

The tissue-mimicking phantom used in all of the experiments was polyacrylamide gel containing bovine serum albumin (BSA).[15,29] This optically transparent gel tissue phantom has acoustic and thermal properties similar to tissue, although the acoustic attenuation is lower, about one-third of the attenuation in tissue. Advantages of using a tissue phantom instead of tissue include the repeatability and uniformity of the phantom's acoustic, thermal, and magnetic properties. Samples were degassed in a desiccant chamber for 1 h prior to polymerization. The axial distance from the transducer face to the proximal end of the sample was 35 mm. The geometrical focus of the transducer was within the sample, 9 mm from its proximal end.

Before the MRI experiment, a fiber-optic probe hydrophone (FOPH 2000, RP Acoustics, Leutenbach, Germany) with 100-μm active diameter was used to measure the focal pressure waveform in water and in the gel for the chosen 63-W source output. Waveforms were measured at the spatial maximum of the peak positive pressure, which coincided with the geometric focus of the source and was 44 mm from the transducer in water and in the water/gel path. Measurements in water were performed with and without the cylindrical housing attached to ensure that the housing did not alter the waveform through reflection or scattering. Measurements in gel were performed without the housing in a slightly different experimental arrangement, but with the same propagation path in gel as used in MRI experiments. The focal waveforms were also modeled in water and in gel using source parameters (aperture, curvature, and electroacoustic efficiency) determined through previously described calibration of an identical source.[28]

A rf power meter (model 21 A, Sonic Concepts, Woodinville, WA) was used to monitor the electrical power delivered to the transducer. The power meter readings were recorded by a digital acquisition (DAQ) board (model 6062E, National Instruments, Austin, TX) at 1 kHz. Fluctuation in the power meter signal was used as an indicator of boiling as has been reported previously.[19,30] Fluctuation in the power delivered to the source is the result of variations in the acoustic impedance, caused by the ultrasound that is backscattered from bubbles. The fluctuation is the most pronounced when bubbles appear at the transducer focus. Before the experiments in the magnet, this system was tested simultaneously with other indicators of boiling and cavitation, including a 20-MHz passive acoustic detection[9,16,19] and a high-speed video camera.[31] Obvious fluctuation in the power meter signal was observed only with boiling and not with cavitation. Passive acoustic detection, however, did detect cavitation and was also used to measure the acoustic pressure threshold for cavitation in the tissue phantom. Time to boil, yielded by all the measuring techniques, agreed within a few milliseconds (roughly the camera frame period and the DAQ sampling period). In replicate samples, boiling occurred in 7.2 ± 0.3 s. The slight variation was attributed to the possible small difference among the sample batches, initial temperature, and the distance between the sample face and the transducer as well as stochastic variability in the distribution of boiling nuclei.[31] The water temperature in the bench-top experiments was 22 ± 1 °C. In the magnet, the initial temperature of each sample was measured using a thermocouple 2 min before HIFU exposure and was 21 ± 1 °C.

In another set of preliminary experiments, bare-wire thermocouples (130 μm, type E, Omega Engineering, Stamford, CT) were implanted in gels at the focus of the transducer and exposed to the treatment conditions along with simultaneous high-speed camera observation to determine the temperature when boiling occurred. Temperature measurements were recorded at a sampling rate of 250 Hz using a data acquisition device (HP34970A, Hewlett-Packard Corp., Palo Alto, CA). A mount was used to cast the thermocouple and the FOPH in the gel 1 mm apart in the focal plane. Alignment of the HIFU focus with the thermocouple was performed by first finding the peak pressure with the FOPH and then by moving the transducer 1 mm to the thermocouple.

MRI experiments were performed in a 4.7-T Bruker magnet (Bruker Medical Systems, Karlsruhe, Germany) with a 30-cm-diameter bore equipped with a Varian (Varian Inc., Palo Alto, CA) INOVA spectrometer and a custom-built, Alderman-Grant type, rf volume coil with an inner diameter of 2.5 cm. All of the MRI data were collected using a gradient-echo sequence. Both magnitude and phase of the MR signal were acquired to obtain primarily axial images, i.e., images in the plane containing the axis of the cylindrical magnet.

In three of the experimental samples, images were collected during very low energy ultrasound exposure (<2 °C temperature rise for <2 s) 10 min prior to experimental exposure in order to confirm the location of the HIFU focus relative to the center of the magnet and, therefore, to define the location of the acquisition volume. This step was deemed unnecessary for the other three samples, and no difference was observed between the two groups. Single-slice gradient-echo MR images were then acquired continuously for 24 s before, 20 s during, and 100 s after HIFU exposure. The following acquisition parameters were used: matrix size 64×128 pixels, TE=4.2 ms, TR=20 ms, flip angle=20 deg, total image acquisition time of 1.3 s, and field of view (FOV) 30×40 mm^2. The in-plane resolution was 0.3×0.5 mm^2 and the slice thickness was 2 mm. Several minutes after HIFU exposure, high spatial resolution axial and transverse MR images were acquired to locate and resolve residual

bubbles. In the axial sequences, the following parameters were used: 256×256 pixels, TE=7.4 ms, TR=300 ms, flip angle=20 deg, FOV=30×40 mm^2, and slice thickness was 2 mm. The voxel volume in these high-resolution images was 120 μm×160 μm×2 mm.

The gradient-echo data were reconstructed to generate both magnitude and phase images. The magnitude images were used to visualize bubbles, and the phase images were used to determine the proton resonance frequency shift.[5] Phase shifts were converted to temperature changes using a calibration curve that was obtained for heated water prior to the HIFU measurements.[32] The calibration measurement was performed in water, not in the tissue phantom, which is 95% water, to ensure the most uniform temperature throughout the sample volume. To obtain the calibration curve, hot (initially boiling) water was poured into a 6-mm-diameter plastic tube centered in the rf coil used in the HIFU experiments. The tube was plugged to ensure that there was no water flow. A thermocouple was positioned on the tube axis, 2 cm away from the coil center. Gradient-echo images (TR/TE/flip angle=24 ms/4.1 ms/20 deg, matrix size 64×32, FOV 2×2 cm^2, and slice thickness 2 mm) were acquired to monitor resonance frequency changes as the water cooled from 90 to 40 °C.

Eddy currents within the gel, generated by the rapidly switching gradients, caused transient, repeatable phase shifts to occur at the onset of MR data acquisition in the absence of temperature changes. These phase shifts could be mistaken for substantial temperature changes. To correct for this potential artifact, a series of control images was acquired from each gel prior to heating. The MR parameters were identical to those described for the gradient-echo images above, and the resulting series of phase images was subtracted, image by image, from those acquired during HIFU exposure.

IV. RESULTS

A. Measurements and calculations to characterize experimental conditions

An experimentally measured calibration curve was obtained to determine temperature changes from the MR acquired phase shifts in the tissue phantom. In Fig. 2, the calibration curve of the temperature versus the absolute value of the water proton chemical shift is plotted. A third-order polynomial was determined based on a least-squares fit of the experimental data points:

$$\delta T = 0.000\,947\,87\delta\varphi^3 - 0.004\delta\varphi^2 + 0.7742\delta\varphi, \qquad (9)$$

where $\delta\varphi$ is the phase shift of the MR signal and δT is the temperature change relative to 40 °C.[32] Equation (9) was used to obtain temperature maps during HIFU exposure instead of the linear relationship 0.009 ppm/°C that is commonly used for lower field magnets and smaller temperature variations.[33,34] The calibration measurement was only performed one time because it was very difficult safely to get a sufficient volume of water uniformly heated to near 100 °C into the bore of the magnet. Hence, a degree of skepticism should be applied when interpreting the results in Fig. 2. While there are evidence in literature, including Ref. 37

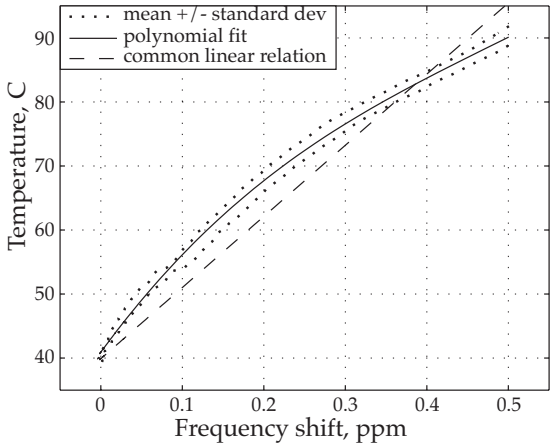


FIG. 2. Water proton resonance frequency shift calibration curve obtained in the 4.7-T magnet that was used for MR thermometry in present work. The solid line shows a third-order polynomial fit to the mean temperature as determined from 12 measurement points located near the center of the rf coil. The dotted lines show the mean ± standard deviation of the measurements. The linear conversion curve (dashed line) often used in 1.5-T magnets is also shown for comparison.

commonly cited for the linear relation, that the dependence between proton resonance frequency and temperature is nonlinear,[35–38] the exact nature of the dependency over broad temperature ranges at 4.7 T requires further validation. Nevertheless, note that the use of the linear relationship would yield *lower* peak MR-determined temperatures than those reported here.

Focal waveforms were measured and modeled in water [Fig. 3(a)] and in the tissue phantom [Fig. 3(b)]. The excellent agreement between the waveforms measured in water with and without the cylindrical sample housing attached to the transducer (Fig. 1) indicates that the housing in the MRI experiments did not alter the acoustic field. Modeling of the field could thus be performed assuming free field conditions. The modeled waveforms also agree well with those measured experimentally both in water and in the phantom [Figs. 3(a) and 3(b)], which supports the use of simulations to determine acoustic and temperature distributions in the phantom. Note also that part of the calibration procedure[28] is to attain agreement between model and measurement in the acoustic distribution for low acoustic excitation. Figure 3 shows that for the source power used in experiments, the combined effects of nonlinear propagation and diffraction led to a typical asymmetric and distorted focal waveform.[12] The compression phase was higher in amplitude, steeper, and shorter in duration than in the linearly predicted waveform (shown as a dashed curve). The peak positive pressure was 30 MPa in water and 27.5 MPa in the phantom. The rarefaction phase was lower in amplitude, smoother, and longer in duration than in the linear waveform with peak negative pressure of 8.4 MPa in water and 8.6 MPa in the phantom. Since nonlinear distortion was clearly observed in the focal waveform, enhanced heating due to nonlinear acoustic propagation effects was expected.

Figure 4 summarizes the results of linear (dashed curves) and nonlinear (solid curves) simulations for various acoustic field parameters and temperature rise in the phan-

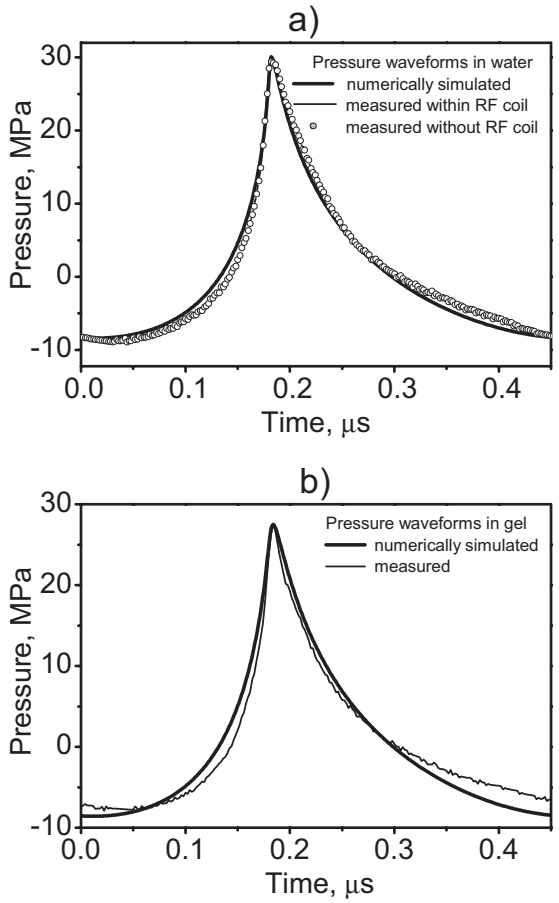


FIG. 3. Focal waveforms simulated numerically and measured by the fiber-optic hydrophone in water with and without the cylindrical, water-filled chamber attached (a) and in the phantom (b). The cylindrical housing did not alter the waveform, and the numerical data are in excellent agreement with the measurements.

tom. Distributions of peak pressure (a), acoustic intensity (b), heat deposition rate (c), and temperature after a 7-s exposure (d) are presented on the HIFU axis (top row) and across the axis in the focal plane $z=44$ mm (bottom row). The −6-dB pressure beam width and length calculated linearly were 1 mm and 6 mm. Nonlinear effects result in reduction in the focal dimensions for the peak positive pressure (0.5 mm and 4 mm) and enlargement of the dimensions for the peak negative pressure (1.2 mm and 8 mm). The maximum value of the peak positive pressure doubled, and the peak negative pressure dropped 30% from the linear to the nonlinear waveform. The *in situ* spatial peak intensity $I=5670$ W/cm^2, Eq. (4), is slightly higher than the linearly calculated, Eq. (6), intensity $I_L=4830$ W/cm^2. The heat deposition rate at the focus is twice that calculated linearly. However, since the heating is over a long time (in seconds) and the extra nonlinear heating is so localized in space, diffusion smoothes the temperature distributions over time, and the final peak temperatures differ much less between the linear and nonlinear simulations than do the peak heating rates. Nonlinear simulations indicate a peak temperature of 99 °C, which we will refer to hereafter as 100 °C because it is so close, while linear simulations yield 86 °C after 7 s of exposure. Dashed vertical lines in Figs. 4(c) and 4(d) show the slice thickness, 2 mm, of the MR voxel. The other voxel dimension in the focal plane is 0.5 mm, and the voxel length in the axial direction is 0.3 mm. The dimensions of the voxel are comparable to those of the heated region. In particular, the 2-mm thickness of the voxel is larger than the half maximum beam width of both the heat sources and even the radial temperature rise profile. Spatial averaging over one voxel can, therefore, be of importance in MR temperature measurements.

At the acoustic exposure levels used in the MR experiments (Fig. 4), cavitation was observed immediately (within a few acoustic cycles) in the phantom with a 20-MHz passive cavitation detector (PCD); in other words, cavitation was present throughout all of the MRI experiments. Figure 5(a) shows the measured threshold curve for cavitation in the phantom for 5-s exposures. For *in situ* peak negative pressures larger than 5 MPa, cavitation was measured immediately in all samples. As shown above, the peak negative pressure at the focus in this study was 8.6 MPa (Figs. 3 and 4)

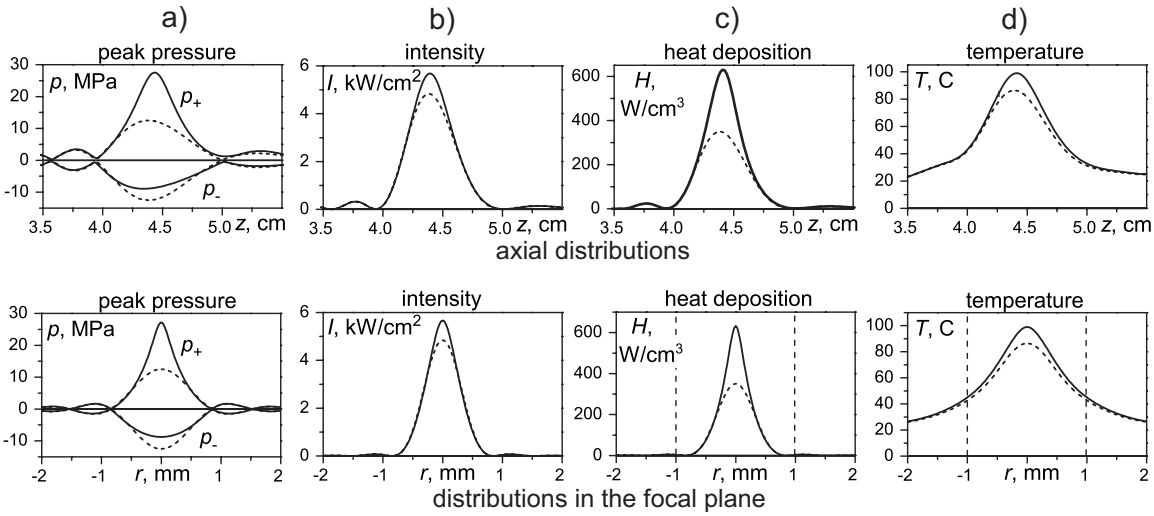


FIG. 4. Simulation results for acoustic and temperature fields in the phantom obtained assuming nonlinear (solid) and linear (dashed) acoustic propagations. Spatial distributions of the peak positive and peak negative pressures, intensity, heat deposition rate, and temperature after 7 s of HIFU exposure are presented axially (upper row) and in the focal plane radially (lower row). Dashed vertical lines on the heat deposition and temperature plots indicate the width of the voxel.

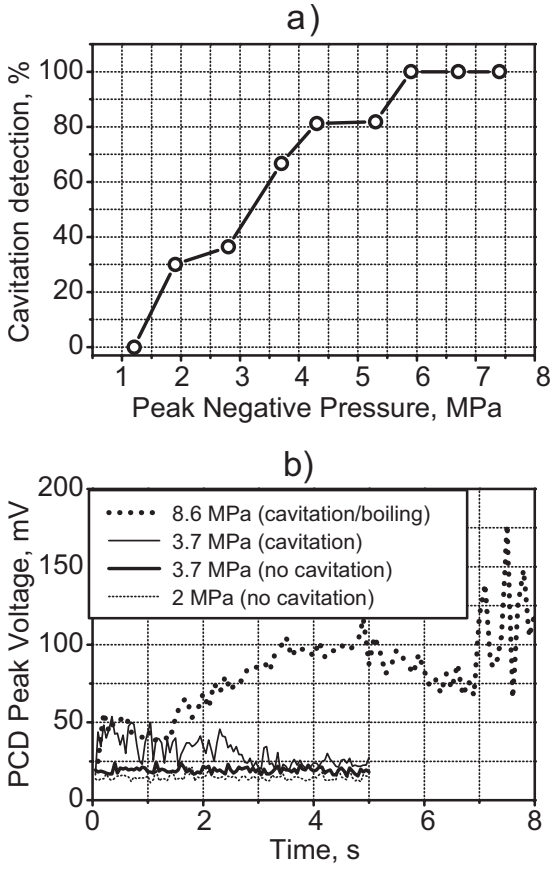


FIG. 5. (a) Percent of times in seven 5-s exposures that cavitation was detected versus peak negative pressure of the HIFU exposure. Cavitation was detected with a 20-MHz PCD high-pass filtered at 15 MHz. Peak negative pressures larger than 8 MPa were used in this work; therefore, cavitation was present in all experiments. (b) Peak-detected representation of time-domain trace recorded by the PCD for three HIFU exposure levels. At the lowest exposure (2-MPa negative pressure), the signal was at the noise level, which was 20 mV. An increase to 3.7-MPa negative pressure caused little change in one case and significant increase in signal amplitude in the other. The large increase in signal was attributed to broadband emissions from cavitation. Under the exposure used in the MR experiments (8.6-MPa negative pressure), the elevated signal due to cavitation was observed immediately after HIFU was turned on, and the signal further increased at 7 s when boiling occurred.

and was, therefore, above the cavitation threshold of the tissue phantom. In an effort to simplify the display of the cavitation signals in Fig. 5(b), time-domain traces measured by the PCD were broken into 100-μs segments, and the peak value of each segment is shown. The background noise level was 20 mV. At an output of 2-MPa peak negative pressure, the signal was only slightly above the noise level of the detection system and was, therefore, interpreted as not having cavitation activity present. Separate exposures at a peak negative pressure of 3.7 MPa created two distinct curves. The lower one was only slightly higher than the curve for the 2-MPa negative pressure exposure, and again the interpretation was that no cavitation occurred. Increasing the applied pressure had little effect on the amplitude of the detected signal without cavitation. However, when cavitation was present as in the upper curve for the 3.7-MPa negative pressure exposure, the signal level from time zero is significantly elevated. Exposure to the conditions described in Fig. 4 and used in the MR experiments here produced the upper curve

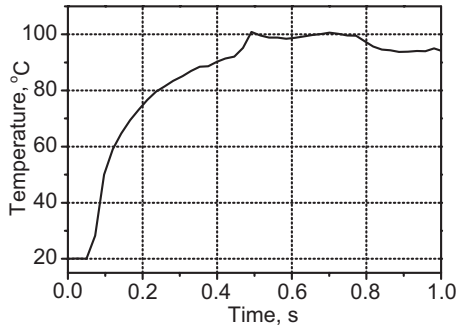


FIG. 6. Temperature measured by a thermocouple at the focus. At the point boiling was visually observed in simultaneous high-speed camera images, 100 °C was measured, and the temperature rise suddenly plateaued. However, the presence of the thermocouple accelerated heating, and therefore, boiling occurred at 0.5 s with the thermocouple present compared to 7 s without it.

labeled 8.6 MPa. An elevated signal due to cavitation is seen immediately at 0 s after HIFU was turned on. The signal is significantly increased when boiling starts at 7 s and is, therefore, distinguishable from the cavitation signal.

The results of temperature measurements with a thermocouple at the focus under the HIFU exposure conditions stated above are shown in Fig. 6. At the instant a boiling bubble appeared, 100 °C was recorded. However, clearly the measurement tool, the thermocouple, affected the measurement: boiling occurred on the thermocouple, and boiling occurred in 7 s without the thermocouple present and in 0.5 s with the thermocouple. Thermocouples are known to create additional heating, near the thermocouple surface, from the viscous damping of the ultrasound-induced motion of the tissue phantom relative to the thermocouple.[39,40] This heating can be reduced by placing the thermocouple at a pressure null instead of at the focus; however, it is evident from Fig. 4(d) that in this case the peak temperature could not be revealed since the linear and nonlinear simulations show the same temperature at the null but significantly different peak temperatures. Viscous heating of the thermocouple is likely responsible for the accelerated heating measured at the focus in the presence of the thermocouple (Fig. 6); nevertheless, when boiling inception was observed by high-speed camera, which visually appeared similar with and without the thermocouple present, the temperature was 100 °C.

B. MRI measurements of HIFU heating in tissue phantom

The following data were all collected in the MRI system from a single exposure of a single sample. The results were repeated in five other samples. The standard deviation in the time to boil between different samples was 0.2 s and was attributed to differences in the sample length, initial temperature, and gas content, as well as variations in the distribution of bubble nuclei.

Power fluctuation was detected by the power meter, suggesting the presence of boiling, after 7.1 s of HIFU heating (Fig. 7). The power fluctuation continued to the end of HIFU exposure. Evidence of boiling was also observed in the amplitude MR images beginning from 7 s (Fig. 8). During

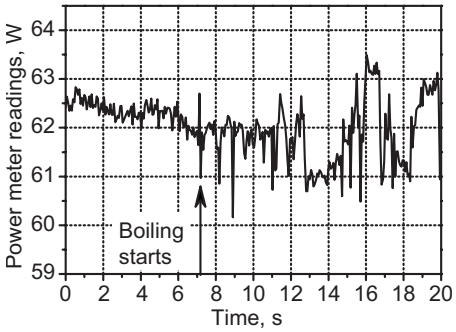


FIG. 7. Indication of boiling after 7-s exposure by fluctuation of the electrical power delivered to the transducer in the MRI experiment.

HIFU exposure, before boiling [Fig. 8(a)], a dark region was seen at the transducer focus. We speculate that this dark region was due to heating, which shifted the resonance frequency away from the bandwidth of the rf excitation pulses and altered the T1 and T2 relaxation times. The first image frame obtained within the time interval of 6.5–7.8 s, following the observed fluctuation in the power meter, contains what looks like bubbles. There is at least one mm-sized, circular, dark region in the image. The dark region indicates a loss of MR signal and would be expected if air or water vapor were present. There is also white and dark mottled "ghosting" around the circular cores. This motion artifact would be expected if the bubbles were oscillating and displacing nearby protons. Figure 8(c), taken 14.3 s after the start of HIFU exposure, shows a dark region of heated tissue containing a bubble. The motion artifact is not seen in this image, possibly because bubble collapses may become less violent over time. For several minutes following treatment, the slowly dissolving bubble or bubbles persisted in the gel sample and were visible to the naked eye and in the MR images. We believe that the reason for bubble persistence is that the samples were incompletely degassed and outgassing into the initially vapor-filled bubbles took place. Figure 8(d) shows a high-resolution image taken 2 min after HIFU exposure when the sample had cooled. Evidence of a bubble is seen within a few voxels of the hottest voxel in the sound field.

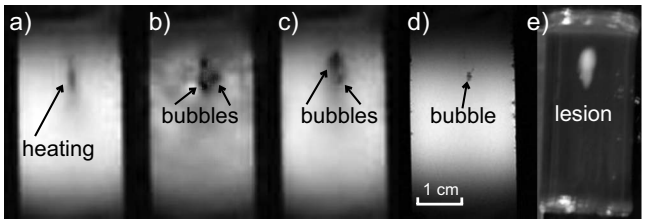


FIG. 8. MR magnitude images of the tissue-mimicking phantom during and after HIFU exposure. The transducer was located above the top of the images, and the sample was exposed for 20 s. (a) The image taken 5.2 s after the start of the exposure shows evidence of heating in the focal region, but boiling had not yet occurred. No evidence of cavitation is observed in the image. (b) After 7.8 s, the image shows large boiling bubbles surrounded by motion artifact. (c) After 14.3 s, motion artifacts are less evident as bubble motion may have become less violent. (d) Even 2 minutes after HIFU exposure, a high-resolution image shows the residual bubble at the HIFU focus. The bubble position corresponded to the distal end of the region of thermally denatured protein (the lesion) photographed in (e). The lesion had grown and enlarged in the direction of the transducer as has been reported to be caused by the presence of boiling (Ref. 15).

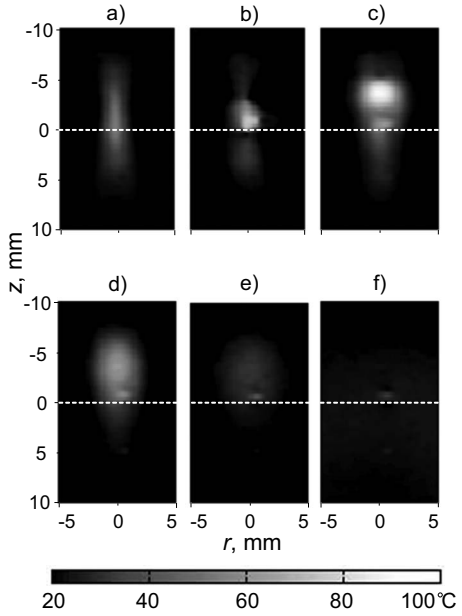


FIG. 9. Two-dimensional temperature distributions measured by MR thermometry: (a) 6.4 s, (b) 7.7 s, and (c) 12.8 s after HIFU was turned on and (d) 3.4 s, (e) 20.3 s, and (f) 40 s after HIFU was turned off. Before boiling (a), the region grew nearly symmetrically about the focus. No temperature field distortion was observed even though cavitation was present. After boiling occurred at 7.1 s, the heated region migrated upward toward the HIFU source and broadened.

The lesion—the region of thermally denatured protein—is not observed in the MR images in Figs. 8(a)–8(d), which is typical with the non-contrast gradient-echo sequence magnitude imaging method employed here.[5] But the final lesion shape, shown in the photograph in Fig. 8(e), indicates that boiling occurred. The lesion continued to be exposed to HIFU for 13 s following the start of boiling, and therefore the lesion grew toward the transducer and broadened on its proximal side arguably because of backscattering from the boiling bubble.[15]

The MRI temperature map in Fig. 9 shows the heated region migrating and broadening in the direction of the transducer. Figure 9(a) shows the MR temperature map acquired just before boiling started. The heated region was nearly symmetric about the focus. However, after boiling started, the hot spot migrated and broadened in the direction of the transducer, as seen in Figs. 9(b) and 9(c). Following treatment, the region cooled [Figs. 9(d)–9(f)], and all that remained was evidence of the bubble (that could also be seen in Fig. 8) at the focus. This bubble appears as a region of slightly elevated temperature on the MR temperature map; however, it had undoubtedly cooled to the ambient temperature. The phase difference that appeared as elevated temperature was likely due to the difference in magnetic susceptibilities of the gel and the gas in the bubble.[37] Both this susceptibility artifact and the motion artifact discussed above mean that MR temperature measurements contain errors once bubbles form and these errors persist as long as the bubbles last.

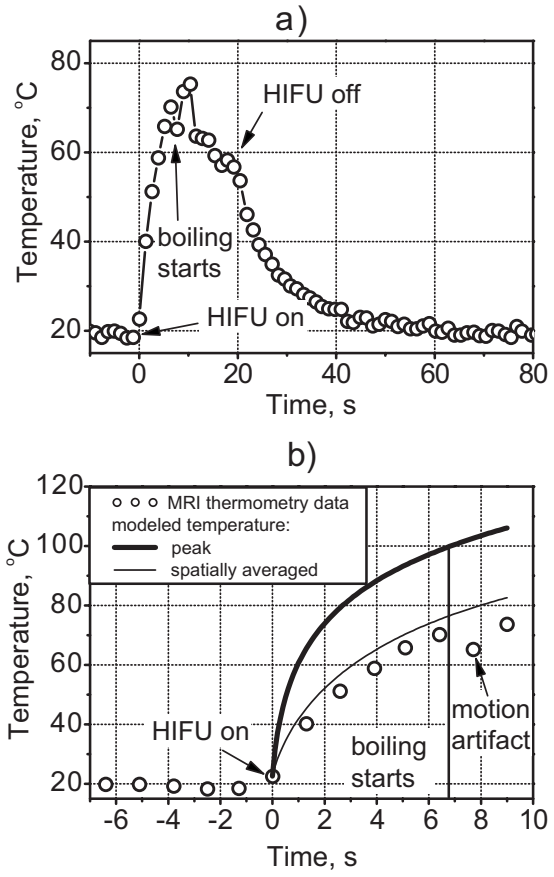


FIG. 10. MR-measured temperature at the focus of the transducer over the course of the treatment (a) and comparison of measurement and calculation in the pre-boiling part of the curve (b). MRI generally tracked the heating during and the cooling following HIFU. However, because of the presence of boiling bubbles during HIFU exposure from 7 to 20 s, temperature readings were erratic. Immediately before boiling (7 s), the calculated peak temperature was 100 °C. The temperature measured with MRI in the focal voxel and the calculated temperature averaged over the voxel volume was only 73 °C.

Boiling started at 7.1 s of exposure as was indicated by fluctuation in the power to the source and the appearance of what look like bubbles in the MR magnitude image; however, immediately before boiling the MR-measured temperature did not reach 100 °C. Figure 10(a) depicts the temperature versus time of the voxel that corresponded to the focus of the transducer. This voxel had the highest temperature in the field at each time point before boiling. Time $t=0$ corresponded to the beginning of the HIFU exposure. For the first 6.5 s, the temperature rose smoothly to 73 °C, which was repeated to within 5 °C in the replicate samples. At 7 s, the bubble appeared in the location of this voxel, and the temperature readings became erratic. The erratic fluctuation was greatest during HIFU exposure when bubble-induced motion artifacts were present. Some of the temperature fluctuation may also have been due to the moving bubbles deflecting the ultrasound and causing local cooling and heating.[17,41] This part of the time-temperature curve varied from sample to sample in a random manner. After 20 s, the voxel began to cool and returned to ambient temperature in about the time expected from calculations. Note that there was no residual artificial elevation in temperature reading for this particular voxel. Although the initial bubble did form in the voxel of interest, its violent and rapid growth moved the bubble off the voxel. In other data sets, the bubble remained in the voxel of interest, and the temperature readings returned to about 10 °C above ambient because of the magnetic susceptibility difference between gas and gel.

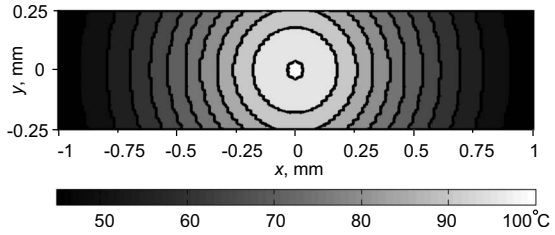


FIG. 11. Temperature distribution after 7-s exposure calculated within a single voxel in the plane perpendicular to the acoustic axis at the spatial peak of temperature. The measured MRI temperature was the temperature averaged over the voxel volume and not the peak temperature.

C. Comparison of MRI measurements and modeling

Unlike the measured temperature of the voxel, the calculated peak temperature did reach 100 °C at 7 s. The thick line in Fig. 10(b) shows the calculated temperature rise using the approach outlined in Sec. II. The circles are the MR-measured data points from Fig. 10(a). However, the temperature field was calculated on a fine grid, and it was necessary to account for the spatial averaging taking place over the finite volume of the MRI voxel (Fig. 4). Figure 11 shows the calculated focal temperature field at 7 s within the voxel cross-section transverse to the acoustic axis. The calculated temperature field was converted to phase using the calibration curve. Then, all phase values within the volume of a voxel were averaged. Finally, the average value was converted back to temperature and indicated the value recorded from the voxel by the MR system. In this manner, the thin line in Fig. 10(b) was obtained, and agreement with measurement is excellent. Indeed, since the spatial temperature distribution close to the focus is very narrow, after averaging it over the size of one voxel the temperature becomes 27 °C lower.

Thus, when spatial averaging was taken into account, the measured maximum of 73 °C corresponded directly to the calculated 100 °C peak temperature. The measured data are shifted slightly to the right of the calculated line, but it is accurate to add a 1.3-s long error bar to the left of the measurement, corresponding to the MR image acquisition time.

V. DISCUSSION AND CONCLUSIONS

In this paper, MR was used as an imaging technique to observe HIFU-induced boiling in a tissue phantom and simultaneously to monitor the temperature as boiling was approached. Standard techniques for MR thermometry were used, but the experiment was designed such that the MR temporal and spatial resolution was better than likely to be achieved clinically. The MR measurements were performed under well-controlled experimental conditions. Homogeneous phantoms with repeatable acoustic and thermal properties were used, the pressure and temperature fields produced by the HIFU source were characterized, the cavitation

threshold in the phantom was measured, and MRI temperature measurements were calibrated prior to the experiment. In other words, the timing and appearance of boiling and of cavitation were well established in advance of testing MR magnitude imaging and thermometry.

Boiling was detected in the MR magnitude images after 7 s of HIFU exposure as circular dark regions surrounded by mottled ghosting indicative of a motion artifact. MRI-detected initiation of boiling correlated with simultaneous observation of fluctuation in the power to the HIFU transducer due to reflections from the bubble and a temperature of 100 °C in numerical simulations. The dark regions in the MR images were about the size of visually observed boiling bubbles (~1-mm diameter), and the motion artifact was thought to be due to bubble motion in the HIFU field. The artifact went away when HIFU exposure was ceased, but a dark spot indicating the presence of the bubble remained.

Although cavitation was present from the beginning of HIFU exposure, cavitation, unlike boiling, was not readily observable in either the power meter reading or the MR magnitude image. The MR image showed darkening in the focal region that grew with time prior to the bubble appearance, but this effect was consistent with the formation of a growing heated region shifting the resonance in the region from the imaging frequency and not with cavitation that is arguably more diffuse and sporadic and less likely to grow symmetrically over time. We speculate that cavitation activity was not observed in the MR imaging because the micrometer-sized cavitation bubbles were much smaller than the millimeter-sized boiling bubbles. In addition, there was no evidence that the cavitation caused additional heating in the samples. Cavitation was present in experiments but not included in the model, yet calculations showed 100 °C at the precise time when boiling bubbles were observed by camera and by artifacts in the MR images. Note that by 100 °C we mean that the temperature was within a few degrees of 100 °C. Boiling may occur at 100 °C or within a few degrees of superheat because of stochastic delays in obtaining a nucleus.[31] In our experiments, the time to boil varied ~4% from sample to sample. In addition, the physical properties of the phantom used in simulations are also known to about this resolution. Thermocouple measurements showed a temperature of 100 °C when boiling started.

Although several independent measures detected either boiling or 100 °C at the same time, the MRI thermometry measurements showed only 73 °C when boiling occurred. If a MR phase versus temperature curve had not been measured and a common linear relationship used instead (Fig. 2), temperatures determined by MR thermometry would have been even lower. MRI alone, therefore, underestimated the peak temperature. However, the results of MR thermometry agree very well with simulations when the modeled temperature distribution was averaged over the volume of the MRI voxel. Spatial averaging was, therefore, shown to be an important factor in this MRI experiment, even though the size of the voxel was much smaller than the voxel size used in clinical MRI systems.[1] Conversely, the transducer, the frequency, focal dimensions, and heating volumes fall within the ranges reported for HIFU clinical work.[12] If only MR thermometry were employed, it would be easy to misinterpret boiling as cavitation because the MRI showed a temperature significantly lower than 100 °C. The temperatures measured by MRI before boiling occurred were underestimated by about 30% due to spatial averaging, which implies that MR thermometry would not have yielded reliable estimations of the thermal dose. Higher frequency, longer focal lengths, and higher amplitudes would shrink the focal width and exacerbate the problem, whereas slower heating allows more time for heat diffusion and tissue motion, which creates other problems, to broaden the heated region.

A deeper look at the spatial and temporal resolutions, particularly, the relation between MR image acquisition time and temperature increase rate, reveals further challenges. We acquired 64 lines of *k*-space at a rate of one line every 20 ms. During the 1.3-s acquisition period the temperature—and therefore the resonant frequency—in some voxels changed substantially, by as much as 20 °C at the focal point. We then used a Fourier transform, which assumes stable resonant frequencies throughout the *k*-space acquisition, to convert to image space and determine the resonant frequency shift in each image voxel.[42,43] Although it is potentially flawed, this same general approach is typically implemented to monitor HIFU procedures using state of the art clinical MR systems. While modern clinical magnets, using parallel imaging techniques combined with rapid acquisition schemes, could provide better temporal resolution than we achieved, it would be very challenging to acquire images of human anatomy with sub-millimeter spatial resolution and acquisition times short enough to adequately reduce resonant frequency changes as *k*-space is filled.

It is important to distinguish boiling from cavitation in clinical research, as detection of boiling could be a useful surrogate for a temperature measurement. Its appearance indicates that the temperature has risen close to 100 °C. Anand and Kaczkowski[22] and Khokhlova *et al.*[44] proposed to use the measured time to boil, combined with calculations, to obtain *in situ* pressures and *in situ* heat rates, which enable these authors to control more precisely clinical HIFU exposure. Further, it is also important to know when boiling occurs or preferably when it will occur because once boiling begins, the treatment is dramatically and irreversibly changed as evidenced by the distortion of the lesion shown in Fig. 8. Undetected boiling will result in lesions that are larger in size and possibly in the wrong location. The clinical significance of this work is that MRI-based temperature measurement alone may be insufficient to monitor and control therapy when treatment temperatures reach 100 °C.

Another point of clinical concern is that boiling can occur in HIFU much more quickly than might be expected. The relationship between focal heating and either focal pressure or drive power is highly nonlinear, especially when output levels are sufficient to cause shock waves. The HIFU heating rate used in our experiments was relatively slow in order to accommodate the long MR image acquisition time. The transducer was excited by 63 W of electrical power, which resulted in a focal pressure waveform that was distorted but not yet shocked, and boiling occurred in 7 s. However, if the power is increased, the heating rate increases significantly. In

our previous experiments, an identical transducer (2.158-MHz, 42-mm aperture and 44.4-mm radius of curvature) was driven up to 300 W, shock waves formed at the focus, and boiling was obtained in under 3 ms in a similar tissue phantom.[16] This time is much shorter than the 1.3-s MR slice acquisition time used in the current study.

These results—HIFU initiation of boiling is unlikely to register 100 °C on MR thermometry and high power can result in millisecond boiling—lead us to a new interpretation of the paper by Sokka *et al.*,[10] which is often cited as the *in vivo* evidence for cavitation-enhanced heating by HIFU. Except for the HIFU power, the study reported here was very similar to the one reported by Sokka *et al.* MRI acquisition time in that work was 4.4 s, and the voxel size was $0.8 \times 1.2 \times 3$ mm^3—larger than in our case. In addition, the authors averaged nine voxels to obtain one temperature reading, which further reduced the spatial resolution. The HIFU source was larger but more focused than the one used in our work (aperture 10 cm and radius of curvature 8 cm), with about the same frequency, 1.7 MHz. The tissue path was 1–2 cm as was ours in phantom. Sokka *et al.* compared two exposures that each had the same total energy delivered to the source: the first was a continuous excitation with electrical power of 28 W for 20 s, and the second contained a 300-W, 0.5-s long burst followed by continuous excitation with electrical power of 21 W for 19.5 s. The authors found that the MR-measured temperature was higher for the second type of treatment, particularly because of greater temperature rise during the initial high-amplitude burst, and the resulting lesions were larger and had migrated toward the source. Bubble activity was detected only in the exposure that included the initial high-amplitude pulse and was attributed to cavitation, not boiling, presumably on the argument that 100 °C was not measured by MRI. The conclusions were, "…cavitation can be reliably used to create significantly larger lesions (three times larger lesions than with conventional focused ultrasound) *in vivo*." In light of our results, an alternative interpretation can be proposed. Heating from the incident HIFU waves, which were arguably shocked or at least significantly distorted, accounted for the enhanced temperature rise during the first 0.5 s of exposure, and the detected bubble activity was boiling, not cavitation. A temperature of 100 °C was not detected just before boiling started because the heated region was smaller than the volume of the MR voxels averaged to measure temperature. Furthermore, after boiling started within 0.5 s of the high-amplitude burst exposure, the presence of bubbles distorted MR temperature readings. By this interpretation, the paper of Sokka *et al.* provides evidence of nonlinearly enhanced or shock-wave-enhanced heating and boiling *in vivo*, not cavitation-enhanced heating *in vivo*. Since MR-based temperature measurements are currently used to calculate ablation volumes in MR-guided HIFU systems, it is important that the respective roles of cavitation and boiling be more clearly understood.

ACKNOWLEDGMENTS

We thank our collaborators at the Center for Industrial and Medical Ultrasound specifically Mr. Aaron Midkiff (Department of Electrical Engineering, University of Washington UW) and Mr. Fran Olson (Applied Physics Laboratory, UW) for design of the experimental apparatus. We also thank the reviewers for questions that helped us improve this manuscript. We gratefully acknowledge funding support from NSBRI Grant No. SMS00402/SMST01601, and NIH Grant Nos. DK43881, EB005250, and EB643.

[1] F. A. Jolesz, K. Hynynen, N. McDannold, and C. Tempany, "MR imaging-controlled focused ultrasound ablation: A noninvasive image-guided surgery," Magn. Reson. Imaging Clin. N. Am. **13**, 545–560 (2005).
[2] F. Wu, Z. B. Wang, W. Z. Chen, W. Wang, Y. Gui, M. Zhang, G. Zheng, Y. Zhou, G. Xu, M. Li, C. Zhang, H. Ye, and R. Feng, "Extracorporeal high intensity focused ultrasound ablation in the treatment of 1038 patients with solid carcinomas in China: An overview," Ultrason. Sonochem. **11**, 149–154 (2004).
[3] R. O. Illing, J. E. Kennedy, F. Wu, G. R. ter Haar, A. S. Protheroe, P. J. Friend, F. V. Gleeson, D. W. Cranston, R. R. Phillips, and M. R. Middleton, "The safety and feasibility of extracorporeal high-intensity focused ultrasound (HIFU) for the treatment of liver and kidney tumours in a Western population," Br. J. Cancer **93**, 890–895 (2005).
[4] C. Chaussy and S. Thuroff, "High-intensity focused ultrasound in prostate cancer: Results after 3 years," Mol. Urol. **4**, 179–182 (2000).
[5] N. McDannold, "Quantitative MRI-based temperature mapping based on the proton resonant frequency shift: Review of validation studies," Int. J. Hyperthermia **21**, 533–546 (2005).
[6] S. Sapareto and W. Dewey, "Thermal dose determination in cancer therapy," Int. J. Radiat. Oncol., Biol., Phys. **10**, 787–800 (1984).
[7] K. R. Gorny, N. J. Hangiandreou, G. K. Hesley, B. S. Gostout, K. P. McGee, and J. P. Felmlee, "MR guided focused ultrasound: Technical acceptance measures for a clinical system," Phys. Med. Biol. **51**, 3155–3173 (2006).
[8] F. A. Jolesz and N. McDannold, "Current status and future potential of MRI-guided focused ultrasound surgery," J. Magn. Reson. Imaging **27**, 391–399 (2008).
[9] R. G. Holt and R. A. Roy, "Measurements of bubble-enhanced heating from focused, MHz-frequency ultrasound in a tissue-mimicking material," Ultrasound Med. Biol. **27**, 1399–1412 (2001).
[10] S. D. Sokka, R. King, and K. Hynynen, "MRI-guided gas bubble enhanced ultrasound heating in *in vivo* rabbit thigh," Phys. Med. Biol. **48**, 223–241 (2003).
[11] D. Melodelima, J. Y. Chapelon, Y. Theillère, and D. Cathignol, "Combination of thermal and cavitation effects to generate deep lesions with an endocavitary applicator using a plane transducer: Ex vivo studies," Ultrasound Med. Biol. **30**, 103–111 (2004).
[12] M. R. Bailey, V. A. Khokhlova, O. A. Sapozhnikov, S. G. Kargl, and L. A. Crum, "Physical mechanisms of the therapeutic effect of ultrasound," Acoust. Phys. **49**, 369–388 (2003).
[13] K. Hynynen, "Demonstration of enhanced temperature elevation due to nonlinear propagation of focussed ultrasound in dog's thigh in vivo," Ultrasound Med. Biol. **13**, 85–91 (1987).
[14] D. Dalecki, E. L. Carstensen, and K. J. Parker, "Absorption. of finite amplitude focused ultrasound," J. Acoust. Soc. Am. **89**, 2435–2447 (1991).
[15] V. A. Khokhlova, M. R. Bailey, J. A. Reed, B. W. Cunitz, P. J. Kaczkowski, and L. A. Crum, "The relative role of nonlinear ultrasound propagation and cavitation in acceleration of HIFU therapy," J. Acoust. Soc. Am. **119**, 1834–1848 (2006).
[16] M. S. Canney, M. R. Bailey, V. A. Khokhlova, and L. A. Crum, "Millisecond initiation of boiling by high-intensity focused ultrasound in tissue-mimicking phantoms," J. Acoust. Soc. Am. **120**, 3110 (2006).
[17] C. C. Coussios and R. A. Roy, "Applications of acoustics and cavitation to noninvasive therapy and drug delivery," Annu. Rev. Fluid Mech. **40**, 395–420 (2008).
[18] K. Takegami, Y. Kaneko, T. Watanabe, S. Watanabe, T. Maruyama, Y. Matsumoto, and H. Nagawa, "Heating and coagulation volume obtained with high-intensity focused ultrasound therapy: Comparison of perflutren protein-type A microspheres and MRX-133 in rabbits," Radiology **237**, 132–136 (2005).
[19] C. Thomas, C. Farny, C. Coussios, R. Roy, and R. Holt, "Dynamics and control of cavitation during high-intensity focused ultrasound application," ARLO **6**, 182–187 (2005).

[20]L. A. Crum and J. B. Fowlkes, "Acoustic cavitation generated by microsecond pulses of ultrasound," Nature (London) **319**, 52–54 (1986).
[21]T. D. Mast, V. A. Salgaonkar, C. Karunakaran, J. A. Besse, S. Datta, and C. K. Holland, "Acoustic emissions during 3.1 MHz ultrasound ablation *in vitro*," Ultrasound Med. Biol. **34**, 1434–1448 (2008).
[22]A. Anand and P. J. Kaczkowski, "Monitoring formation of high intensity focused ultrasound (HIFU) induced lesions using backscattered ultrasound," ARLO **5**, 88–94 (2004).
[23]N. A. Watkin, I. H. Rivens, and G. R. ter Haar, "The intensity dependence of the site of maximal energy deposition in focused ultrasound surgery," Ultrasound Med. Biol. **22**, 483–491 (1996).
[24]M. R. Bailey, L. N. Couret, O. A. Sapozhnikov, V. A. Khokhlova, G. ter Haar, S. Vaezy, X. Shi, R. Martin, and L. A. Crum, "Use of overpressure to assess the role of bubbles in focused ultrasound lesion shape *in vitro*," Ultrasound Med. Biol. **27**, 696–708 (2000).
[25]S. Vaezy, X. Shi, R. W. Martin, E. Chi, P. I. Nelson, M. R. Bailey, and L. A. Crum, "Real-time visualization of focused ultrasound therapy," Ultrasound Med. Biol. **27**, 33–42 (2001).
[26]Z. Xu, J. B. Fowlkes, E. D. Rothman, A. M. Levin, and C. A. Cain, "Controlled ultrasound tissue erosion: The role of dynamic interaction between insonation and microbubble activity," J. Acoust. Soc. Am. **117**, 424–435 (2005).
[27]P. P. Lele, "Effects of ultrasound on "solid" mammalian tissues and tumors in vivo," *Ultrasound: Medical Applications, Biological Effects and Hazard Potential* (Plenum, New York, 1986), pp. 275–306.
[28]M. S. Canney, V. A. Khokhlova, M. R. Bailey, O. A. Sapozhnikov, and L. A. Crum, "Acoustic characterization of high intensity focused ultrasound (HIFU) fields: A measurement and modeling approach," J. Acoust. Soc. Am. **124**, 2406–2420 (2008).
[29]C. Lafon, V. Zderic, M. L. Noble, J. C. Yuen, P. J. Kaczkowski, O. A. Sapozhnikov, F. Chavrier, L. A. Crum, and S. Vaezy, "Gel phantom for use in high-intensity focused ultrasound dosimetry," Ultrasound Med. Biol. **31**, 1383–1389 (2005).
[30]L. A. Crum and W. Law, "The relative roles of thermal and nonthermal effects in the use of high intensity focused ultrasound for the treatment of benign prostatic hyperplasia," Proceedings of the 15th International Congress on Acoustics, Trondheim, Norway, 1995.
[31]M. S. Canney, M. R. Bailey, V. A. Khokhlova, W. Kreider, and L. A. Crum, "Observations of cavitation and boiling in a tissue phantom due to high intensity focused ultrasound," J. Acoust. Soc. Am. **122**, 3079 (2007).
[32]K. Marro, D. Lee, T. Khokhlova, and M. Bailey, "Nonlinear frequency shifts during HIFU-induced boiling at 4.7 T," The 48th Experimental NMR Conference, Daytona Beach, FL, April 23–27, 2007.
[33]Y. Ishihara, A. Calderon, H. Watanabe, K. Okamoto, Y. Suzuki, K. Kuroda, and Y.Suzuki, "A precise and fast temperature mapping using water proton chemical shift," Magn. Reson. Med. **34**, 814–823 (1995).
[34]T. Wu, K. R. Kendell, J. P. Felmlee, B. D. Lewis, and R. L. Ehman, "Reliability of water proton chemical shift temperature calibration for focused ultrasound ablation therapy," Med. Phys. **27**, 221–224 (2000).
[35]J. S. Philo and W. M. Fairbank, "Temperature dependence of the diamagnetism of water," J. Chem. Phys. **72**, 4429–4433 (1980).
[36]E. P. Day, "Equation for the magnetic susceptibility of water," J. Chem. Phys. **72**, 4434–4436 (1980).
[37]J. C. Hindman, "Proton resonance shift of water in the gas and liquid states," J. Chem. Phys. **44**, 4582–4592 (1966).
[38]R. Cini and M. Torrini, "Temperature dependence of the magnetic susceptibility of water," J. Chem. Phys. **49**, 2826–2830 (1968).
[39]W. J. Fry and R. B. Fry, "Determination of absolute sound levels and acoustic absorption coefficients by thermocouple probes—Theory," J. Acoust. Soc. Am. **26**, 294–310 (1954).
[40]R. L. Clarke and G. R. ter Haar, "Temperature rise recorded during lesion formation by high intensity focused ultrasound," Ultrasound Med. Biol. **23**, 299–306 (1997).
[41]C. C. Coussios, C. H. Farny, G. R. ter Haar, and R. A. Roy, "Role of acoustic cavitation in the delivery and monitoring of cancer treatment by high-intensity focused ultrasound (HIFU)," Int. J. Hyperthermia **23**, 105–20 (2007).
[42]T., Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy* (Clarendon, Oxford, 1991).
[43]Z.-P. Liang and P. C. Lauterbur, *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*, IEEE Press Series in Biomedical Engineering (IEEE, New York, 2000).
[44]V. A. Khokhlova, M. R. Bailey, J. Reed, M. S. Canney, P. J. Kaczkowski, and L. A. Crum, "Nonlinear mechanisms of lesion formation by high intensity focused ultrasound," AIP Conf. Proc. **829**, 117–121 (2006).

Forward-masking based gain control in odontocete biosonar: An evoked-potential study

Alexander Ya. Supin[a)]
Institute of Ecology and Evolution, Russian Academy of Sciences, 33 Leninsky Prospect, 119071 Moscow, Russia

Paul E. Nachtigall and Marlee Breese
Marine Mammal Research Program, Hawaii Institute of Marine Biology, University of Hawaii, Kaneohe, Hawaii 96744-1106

(Received 6 August 2008; revised 29 January 2009; accepted 5 February 2009)

Auditory evoked potentials (AEPs) were recorded during echolocation in a false killer whale *Pseudorca crassidens*. An electronically synthesized and played-back ("phantom") echo was used. Each electronic echo was triggered by an emitted biosonar pulse. The echo had a spectrum similar to that of the emitted biosonar clicks, and its intensity was proportional to that of the emitted click. The attenuation of the echo relative to the emitted click and its delay was controlled by the experimenter. Four combinations of echo attenuation and delay were tested (−31 dB, 2 ms), (−40 dB, 4 ms), (−49 dB, 8 ms), and (−58 dB, 16 ms); thus, attenuation and delay were associated with a rate of 9 dB of increased attenuation per delay doubling. AEPs related to emitted clicks displayed a regular amplitude dependence on the click level. Echo-related AEPs did not feature amplitude dependence on echo attenuation or emitted click levels, except in a few combinations of the lowest values of these two variables. The results are explained by a hypothesis that partial forward masking of the echoes by the preceding emitted sonar pulses serves as a kind of automatic gain control in the auditory system of echolocating odontocetes. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3089589]

PACS number(s): 43.80.Lb [WWA] Pages: 2432–2442

I. INTRODUCTION

The biosonar of odontocetes (toothed whales, dolphins, and porpoises) operates with echo levels varying by many tens of decibels because of the variation of target strengths and distances. Therefore, the performance of the biosonar requires that its receiving part (the auditory system) be capable of analyzing echoes under widely varying intensities. This capability requires either a kind of automatic gain control (AGC) to decrease the range of variation of intensities of the returning echoes in the auditory system, or a capability of the auditory system to analyze the echo signals over a very wide variety of neuronal response levels. We do not yet know how the odontocete's auditory system solves this problem. One way to keep the echo level within a reasonable range is to vary the level of the outgoing signal. Experiments on animals in the wild have shown that dolphins and whales may vary the intensity of their echolocation pulses according to the distance to the target, roughly at a rate of 20 dB per distance decade (Rasmussen *et al.*, 2002; Au and Benoit-Bird, 2003; Au and Herzing, 2003; Au and Würsig, 2004). To some extent, the variation of the level of the emitted biosonar pulses may compensate for the echo attenuation with distance. However, this mechanism does not control the gain in the receiving part of the biosonar, i.e., in the auditory system. Variation of the emitted pulse intensities does not influence the ratio of neuronal responses to the emitted acoustic pulse and the echo. Since the echo-response may be partially or completely masked by the much more intense preceding emitted pulse, the ratio of responses to the emitted pulse and the echo should also influence the neuronal response to the echo depending on the target strength and distance.

It has been hypothesized that the odontocete auditory system, i.e., the receiving part of the biosonar, possesses an AGC mechanism that is capable of compensating for the distance-dependent variation of responses to the echo. According to this hypothesis, this mechanism is based on partial forward masking of neuronal echo-responses by the more intense preceding emitted acoustic pulses. Originally, the hypothesis was based on results obtained during conditions of *external* double-pulse stimulation instead of natural echolocation. It has been shown that when the evoked response to a test click is forward masked by a more intense conditioned click, the rate of release from this masking was 30–40 dB per delay decade in bottlenose dolphins (Popov and Supin, 1990) or 25–30 dB per delay decade in a false killer whale (Supin *et al.*, 2007). The double-pulse stimulation may be considered as simulating the interaction between the emitted biosonar pulse and the echo in natural echolocation. The first (conditioning) pulse in the pair simulates the emitted pulse and the second (test) pulse simulates the echo. The rates of release from masking found in those experiments were remarkable since they were close to the physical rate of echo attenuation with distance that is between 20 and 40 dB per distance decade depending on the target size. Thus, there was good

[a)]Electronic mail: alex_supin@mail.ru

 0001-4966/2009/125(4)/2432/11/$25.00

reason to hypothesize that if the same forward masking takes place between the emitted pulse and echo, it may make responses to a target invariable of distance. In other words, with distance increase, the decrease in the echo intensity and the release from masking by the emitted pulse compensate one another, thus keeping the auditory response magnitude nearly constant.

There have been some data obtained in favor of this hypothesis. Recordings of brain stem evoked potentials during natural echolocation have shown that the response to the echo remained almost constant with variation of target distance, although the intensity of emitted echolocation pulses varied only slightly (Supin *et al.*, 2004, 2005). These data indicate that AGC in the odontocete biosonar involves a variation of sensitivity of the auditory system. It appears as though the echo-response was kept nearly constant because of its delay-dependent release from forward masking by the previous louder emitted pulse, as predicted by the release from forward-masking hypothesis.

The data on the rate of release from masking obtained with external double-pulse stimulation cannot be directly applied to conditions of natural echolocation, particularly because the rate of release from masking depends on the relation of intensities and frequency bands of the two pulses (Popov *et al.*, 2001). The measurement of the rate of release from masking with conditions closer to real echolocation would be a much more trustworthy confirmation of the hypothesis. Some confirmation has been obtained. Psychophysical measurements in a false killer whale (Supin *et al.*, 2008) have shown that thresholds of detection of simulated electronic echo decrease with the echo delay increase with a rate of −31.5 dB per delay decade. This threshold decrease was interpreted as release from forward masking and the obtained rate of release was not substantially different from that (25–30 dB) previously obtained in passively presented double-pulse evoked-potential measurements and predicted by the forward-masking AGC hypothesis. Note that depending on the target features, the echo level varies with distance at a rate from 20 dB/decade (a flat mirror) to 40 dB/decade (a small object), i.e., 30 ± 10 dB. If the masked threshold depends on distance at a rate near −30 dB/decade and the echo sound pressure level depends on distance at nearly the same rate, then the echo sensation level (relative to the masked threshold) should not be dependent on distance.

However, threshold data do not always allow the prediction of variation in supra-threshold response magnitude because of a possible non-linearity in the auditory system. Therefore, the goal of the present study was to assess the possibility of forward-masking based AGC in the odontocete sonar at supra-threshold echo levels.

We approached the problem by examining the brain auditory evoked potentials (AEPs) during echolocation in an odontocete subject. This approach has demonstrated its potential for the investigation of the odontocete biosonar beginning with the first attempts made by Bullock and Ridgway (1972), which showed that voluntary click emission by a dolphin as well as external stimulation ("artificial echo") resulted in AEPs. The quantitative investigation of outgoing pulse-related and echo-related AEPs has been accomplished in a variety of recent studies using a non-invasive recording technique (Supin *et al.*, 2003, 2004, 2005).

Unlike our previous studies, in the current work, we used a simulated electronic (phantom) echo technique to produce echoes rather than using real echolocation targets (Aubauer and Au, 1998). Using this technique, the echo is produced by electronically playing back an acoustic pulse that is triggered by the animal's emitted biosonar pulse. An advantage of this technique is the possibility of flexibly and independently varying the parameters of the echoes, which is not always possible with real physical targets. By using this technique, we investigated how the echo-produced AEP magnitude varied when the echo delay and level varied in the same manner as expected for a real target at different distances.

II. METHODS

A. Subject and experimental conditions

The experiments were carried out at the facilities of the Hawaii Institute of Marine Biology, Marine Mammal Research Program. The subject was Kina, a false killer whale *Pseudorca crassidens*, an experienced approximately 30–35 year-old female kept in a wire-net enclosure in Kaneohe Bay, HI. This same whale was used in previous evoked-potential studies of echolocation (Supin *et al.*, 2003, 2004, 2005, 2007, 2008; Nachtigall and Supin, 2008). The animal was trained to echolocate, to recognize electronic echoes, and to report the echo presence or absence using a go/no-go reporting paradigm. The subject had significant hearing loss at frequencies above 35–40 kHz; however, its hearing sensitivity below 30 kHz was satisfactory (Yuen *et al.*, 2005).

The experimental facilities were laid out as follows (Fig. 1). The experimental enclosure was constructed of a floating pen frame (1), 8×10 m in size, supported by floats and bearing an enclosing wire net. This enclosure (the animal section) linked to a target presentation section—another floating frame (2), 6×8 m in size that served as a platform for mounting targets or transducers and did not bear net. In the net divider separating these two sections, there was an opening bounded by a hoop (3), 55 cm in diameter that served as a hoop station for the animal. In front of the hoop, a hydrophone (4) was positioned 1 m from the animal's blowhole to record the echolocation pulses. At a distance of 2 m, a transducer (5) was positioned to play-back electronic echoes. The hoop station (3), the hydrophone (4), and the transducer (5) were in a longitudinal straight line; altogether at a depth of 80 cm. In front of the animal, there was a movable baffle (6). When pulled up, this baffle screened the hydrophone and play-back transducer from the animal positioned in the hoop station; when it was lowered down, it opened the space in front of the animal. Near the hoop station, a response ball (7) was mounted above the water surface serving as a target or echo-present response indicator. The animal's position in the stationing hoop was monitored through an underwater video camera (8). The electronic equipment and experimenter were housed in a shack (9) with

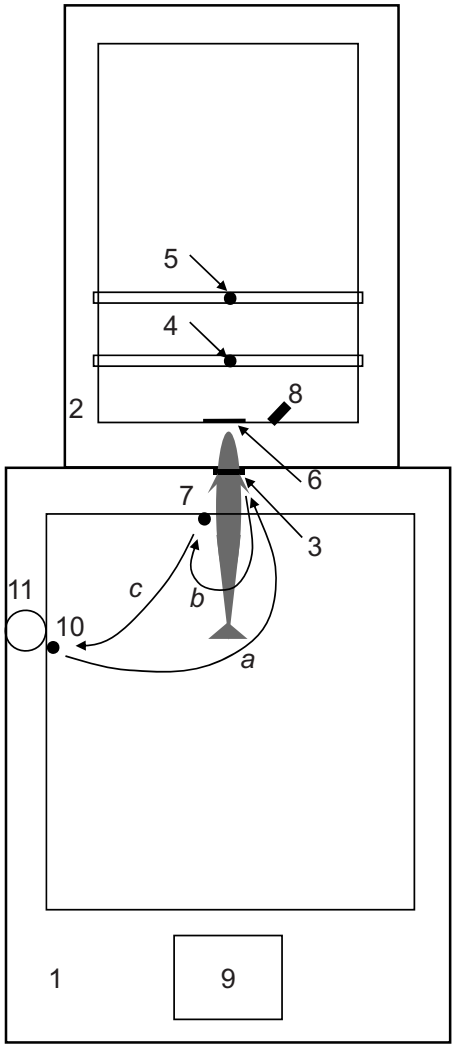


FIG. 1. Experimental design. (1) Experimental pen, (2) transducer-holding frame, (3) hoop station, (4) receiving hydrophone, (5) playing-back transducer, (6) baffle, (7) response ball, (8) video camera, (9) shack, (10) start-position ball, and (11) trainer's position.

a window facing the pen. A ball (10) mounted in the water at the side pen deck nearly the trainer's position (11) marked the start position for the animal.

B. Experimental procedure

Each session consisted of 50 trials randomly alternating: 20 trials of electronic echo play back, 15 trials of real physical target presence (a hollow aluminum cylinder, 38.1 mm outer diameter, 25.4 mm inner diameter, 180 mm long, −22 dB target strength, and 3 m distance), and 15 trials of neither electronic echo nor real target (echo absence). Electronic echo trials served to collect AEP data. Real target trials served to maintain the skill of the subject to detect and discriminate targets, which was necessary for other tasks. Echo-absent trials served to maintain correct echolocation activity reporting echo or target presence or absence. The experimental procedure was as follows.

(i) Each session began with the trainer attaching suction-cup electrodes for AEP recording (see below for detail). (ii) The animal was sent to a hoop station (trace *a* in Fig. 1); the animal swam into the hoop in such a position that the hoop was at the level of the pectoral fins. During the animal's positioning, the baffle (6 in Fig. 1) was in the raised position that prevented the animal from echolocation. (iii) When the animal took the proper position in the hoop station, the baffle was lowered for 3 s, thus allowing the animal to inspect the space by echolocation. The animal immediately emitted a train of echolocation clicks, normally 10–40 clicks in a train. (iv) When either the electronic echo or real target was presented, the animal was required to signal its detection by leaving the hoop and touching a signal ball (trace *b* in Fig. 1), then coming to the trainer for the fish reward (trace *c*). During the no-echo, no-target trials, the animal was required to wait 6 s in the hoop, then it was signaled to leave the hoop (trace *b* was absent) and come for the fish reward (trace *c*).

During the electronic echo trials the electronic echo generation and AEP-acquisition system (see below) were turned on as soon as the baffle was lowered down, and it was kept on until the end of the echolocation click train. Thus, AEPs triggered by echolocation clicks were collected. In real target or echo absent trials, AEP acquisition was turned on for the same time, but the echo generation was turned off.

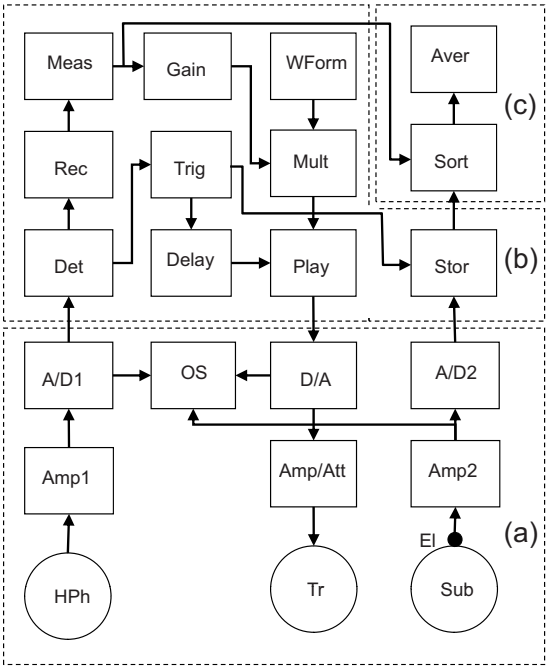


FIG. 2. Instrumentation and data processing. (a) Hardware, (b) on-line software, (c) off-line software. HPh—hydrophone, Amp1—hydrophone amplifier, A/D1—analog-to-digital converter for sound signals, Det—sonar pulse detection, Rec—sonar pulse waveform recording, Meas—sonar pulse amplitude measurements, Gain—echo level control, WForm—echo waveform, Mult—multiplication, Trig—echo and data acquisition triggering, Delay—echo delay control, Play—echo playing back, D/A—digital-to-analog converter, Amp/Att—output amplifier/attenuator, Tr—playing transducer, Sub—subject, El—electrodes, Amp2—brain-potential amplifier, A/D2—analog-to-digital converter for brain-potential signals, Stor—data storage in computer memory, Sort—brain-potential record sorting, Aver—off-line averaging, and OS—oscilloscope.

C. Instrumentation and data collection

The experimental equipment was designed as shown in Fig. 2. It contained two channels, one for recording and pro-

cessing the biosonar sound clicks and another for recording the brain potentials.

The emitted biosonar pulses were picked up by a hydrophone (HPh in Fig. 2) B&K 8103 with the flat frequency response up to 170 kHz. The picked-up signals were amplified by 20–40 dB within a frequency range of 5–100 kHz by a custom-made amplifier (Amp1 in Fig. 2) based on a chip AD820 (Analog Devices). The amplified signal was led to an analog-to-digital converter (A/D1 in Fig. 2) of a data acquisition card NI DAQ-6062 (National Instruments). The digitizing was done at a sampling rate of 512 kHz. The digitized signal was processed by a standard personal computer. The processing included steps as follows.

A sonar click was detected when the input signal exceeded a pre-adjusted threshold level (Det in Fig. 2). The click waveform within a 100-μs window was acquired and stored in computer memory for further off-line analysis (Rec in Fig. 2). The peak-to-peak amplitude of the stored click was measured (Meas in Fig. 2). A pre-synthesized echo waveform (WForm in Fig. 2) was multiplied (Mult in Fig. 2) by the amplitude value of the picked-up click and by a proportionality factor (Gain in Fig. 2); therefore, the amplitude of the synthesized signal was proportional to amplitude of the picked-up sonar click. The proportionality factor was controlled by the experimenter to vary the level of the electronic echo from trial to trial. The obtained signal was played back (Play in Fig. 2) being triggered at the instant that the biosonar click was picked up (Trig in Fig. 2) with a controlled delay (Delay in Fig. 2). The played-back digital signal was digitally-to-analog converted (D/A in Fig. 2) by the same acquisition card NI DAQ-6062, amplified and attenuated (Amp/Att in Fig. 2), and played through an ITC-1032 (International Transducer Corporation) transducer (Tr in Fig. 2).

Brain potentials were picked up by the surface electrodes (El in Fig. 2), which were 10-mm golden-plated disks mounted in 50-mm silicon suction cups. The active electrode was placed at the dorsal head surface midline, 7 cm behind the blowhole; the reference electrode was fixed at the dorsal fin lateral surface. The electrodes were connected by 10-m flexible shielded cables with the input of a custom-made brain-potential amplifier (Amp2 in Fig. 2) providing 20 000 times amplification within a frequency range of 200–5000 Hz. The amplified signal was analog-to-digitally converted (A/D2 in Fig. 2) by the same acquisition card NI DAQ-6062. Fragments of the signal 15–25 ms long (depending on the used electronic echo delay) were extracted and the acquisition of each fragment was triggered by the same trigger (Trig in Fig. 2) as the played-back echoes, however, without delay; thus, the acquired fragment included events in the brain potentials beginning immediately after the biosonar-click detection. The extracted brain-potential fragments were stored in computer memory *without* averaging. Each of the records was marked by the decibel-level of the triggering sonar click.

This cycle of data processing repeated automatically at each picked-up biosonar click. Both input (sound and brain-potential) and output (played-back) signals were monitored by an oscilloscope Tektronix TDS1002 (OS in Fig. 2).

Further analysis of brain potentials was performed off-line. All the brain-potential records were sorted in 5-dB bins by the levels of the triggering clicks. After collection of all the brain-potential records at a certain combination of the echo gain and delay, the records in each bin were averaged, thus obtaining the final records of the brain responses to the emitted pulse and echo, separately for different emitted click levels with a tolerance of ± 2.5 dB. With this procedure, the number of original records was not constant among different averaged records because it depended on the number of biosonar clicks produced by the animal at each particular level. Only those averaged records that were based on not less than 250 original records were accepted as having enough data to produce a satisfactory signal-to-noise ratio.

The biosonar-click data processing (recording the click waveform, measuring its amplitude, and synthesizing the played-back echo waveform) required up to 3–3.5 ms. This resulted in a time limitation for available echo delays. Together with the acoustic delay of 2 ms (sound propagation for 1 m from the animal to the receiving hydrophone plus 2 m from the playing transducer to the animal), the minimal available echo delay could not be less than 5–5.5 ms if the echo was triggered by the same biosonar pulse that was used for calculating the echo parameters. To avoid this limitation, each emitted sonar pulse triggered playing-back the echo waveform computed using the amplitude value of not the current but the *previous* pulse. After triggering, the program computed the signal to be triggered by the next biosonar pulse. With this mode, triggering was possible as soon as a few microseconds after the sonar pulse detection, i.e., with virtually zero delay. Together with the acoustic delay, the low limit of echo delays was 2 ms. This solution was based on an assumption that during a biosonar-click train, the amplitude of the clicks vary rather gradually, not by big jumps from high to low amplitude or vice versa. If so, the error resulting from computation based on previous instead of current sonar pulse would not be large. The validity of this assumption is examined and discussed below based on the click statistical results.

Data processing was controlled by a custom-made program ("virtual instrument") designed on the base of LabVIEW software (National Instruments).

D. AEP magnitude measurements

Amplitudes of both emission-related and echo-related AEPs were measured as peak-to-peak differences within a 3-ms time window around the anticipated response appearance. In doing so, it was kept in mind that the addition of background electrical noise to the response might have increased the peak-to-peak span in the measurement window, thus resulting in overestimate of the response amplitude. To compensate for this effect, the noise power was subtracted from the response power. This was done by subtracting background noise power from an estimate of AEP power using a formula

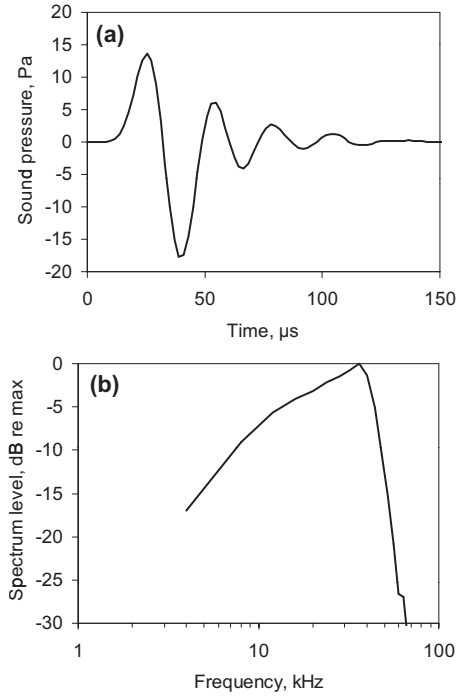


FIG. 3. Electronic echo pulse (150 dB re 1 μPa peak/peak) waveform (a) and frequency spectrum (b).

$$A_r = \sqrt{A_m^2 - (2\sqrt{2}N)^2},$$

where A_m is the measured peak-to-peak amplitude, N is noise rms, A_r is the resulting amplitude estimate, and the factor of $2\sqrt{2}$ relates peak-to-peak to rms values; this value was arbitrarily chosen as valid for a sinusoid. This correction was negligible for responses well exceeding the background noise but reduced the resulting estimates when a measured peak-to-peak span within the measuring window approached the noise level. When the expression under the square root was negative, A_r was adopted as zero.

E. Echo pulse features and variation ranges

Waveform, spectrum, and level of the played-back electronic echo were monitored in the absence of the animal by positioning a receiving hydrophone (B&K8103) in the center of the hoop station.

The spectrum of the acoustic echo pulse was close to those of typical emitted biosonar pulses (see below). The waveform (a) and frequency spectrum (b) of the acoustic echo pulse (150 dB re 1 μPa p/p) are presented in Fig. 3. The echo pulse had a spectrum peak at 35 kHz with high-frequency decay down to −20 dB at 55 kHz and low-frequency decay down to −20 dB at 3 kHz. Note that the spectrum peak fell into a frequency range where the subject had minimal hearing loss.

A distinguishing feature of the experimental design was that the level of electronic echo was not directly controlled by the operator. It was proportional to the amplitude of the emitted biosonar pulses, thus imitating an important feature of the real echo. Instead of the level, the *attenuation* of the echo signal relative to the emitted sonar pulse was controlled by the operator and used to specify the echo. This attenuation was calculated as a decibel-difference between the echo level next to the animal's head and the source level (i.e., level at 1-m distance from the head) of the emitted sonar pulse. With the instrumentation designed as described above, the echo attenuation was determined by (i) the transduction factor of the receiving hydrophone (HPh in Fig. 2), (ii) the transduction factor of the playing-back transducer (Tr in Fig. 2), (iii) the distance from the animal's head to the hydrophone, (iv) the distance from the play-back transducer to the head, (v) the gain of the hydrophone amplifier (Amp1 in Fig. 2), (vi) the gain of the output amplifier-attenuator (Amp/Att I Fig. 2), and (vii) the digital multiplication factor (Gain in Fig. 2). The factors (i)–(iv) were kept constant throughout all the present study, whereas variation of parameters (v)–(vii) allowed echo attenuation variation within a wide range.

Apart from the attenuation, the delay of the electronic echo relative to the instant of the detection of the emitted sonar pulses was varied. Since acquisition of every final record required at least a few experimental sessions, only a limited number of combinations of the echo attenuation and delay were tested in this study, namely,

(1) 2 ms delay and −31 dB attenuation;
(2) 4 ms delay and −40 dB attenuation;
(3) 8 ms delay and −49 dB attenuation; and
(4) 16 ms delay and −58 dB attenuation;

i.e., the echo attenuation and delay were associated with a rate of 9 dB per delay doubling (i.e., 30 dB per delay decade). This rate was chosen because an earlier investigation in the same subject with double-pulse external stimuli had shown that the time of release from forward masking depended on the conditioning-to-test stimuli ratio with just this rate, around 9 dB per delay doubling (Supin *et al.*, 2007). On the other hand, the rate of 9 dB per delay doubling is the midpoint of the range of echo intensity variation depending on distance (from 6 to 12 dB per distance doubling, depending on the target size), so this rate was expected to satisfactorily reproduce the conditions of real echolocation. Assuming the attenuation-vs-delay rate of 9 dB per delay doubling, all the used combinations of attenuation and delay corresponded to a target strength of −26 dB at sound velocity of 1500 m/s.

III. RESULTS

A. Subject performance statistics

In all of the sessions, the subject featured a high percentage of correct responses. The electronic echo was correctly detected in 649 of 670 trials (99.1%), the physical target was correctly detected in 508 of 510 trials (99.6%), and correct rejections were in 500 of 510 trials (98.0%). Only AEP records obtained during correct detection trials were used for further AEP analysis.

B. Emitted sonar clicks' parameters and statistics

During the echo threshold measurements, 21 687 waveforms of emitted sonar clicks were collected and analyzed off line. Representative click waveforms of the most typical

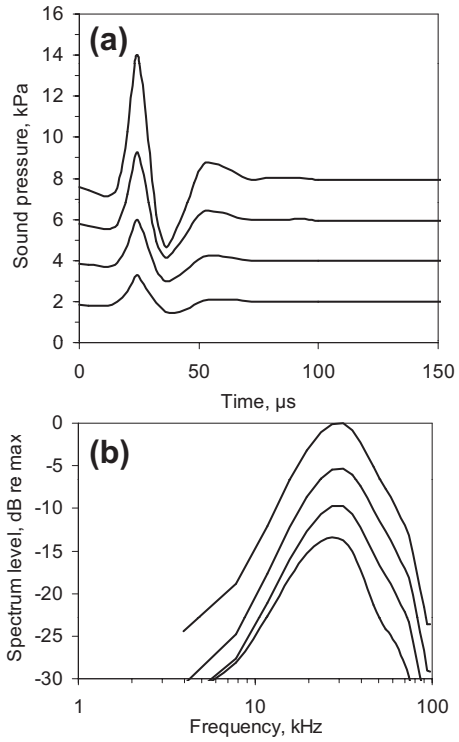


FIG. 4. Emitted sonar pulses of various levels. (a) Waveforms, from top to bottom: 200 ± 2.5, 195 ± 2.5, 190 ± 2.5, and 185 ± 2.5 dB re 1 μPa peak/peak, shifted relative one another by 2 kPa for better discrimination. (b) Corresponding spectra (from higher to lower, respectively).

levels, from 185 ± 2.5 to 200 ± 2.5 dB re 1 μPa peak/peak, are presented in Fig. 4(a), each as an average of 200 examples. The clicks of all levels consisted of a few waves, the main compression wave and decreasing rarefaction-compression oscillations. The frequency spectra of these waveforms had peaks at 27–32 kHz and fell down to −20 dB at 80–88 kHz at the high-frequency branch and at 5–7 kHz at the low-frequency branch. Note that the spectrum peak fell into the frequency range where the subject had minimal hearing loss.

Statistical distributions of click levels are presented in Fig. 5, separately for the four investigated combinations of echo attenuation and delay [(a)–(d)]. From 5367 to 6350 clicks were collected and measured at each of the combinations. All of the distributions demonstrated a rather wide range of click level variation, from 165 to 205 dB re 1 μPa peak/peak; however, more typically click levels ranged from 185 to 200 dB. Mean peak levels noticeably differed among different combinations: from 180.0 dB (at 4 ms, −40 dB) to 191.2 dB (at 16 ms, −58 dB). The means featured a trend: The deeper the echo attenuation, the higher the mean emitted pulse level (Fig. 6). As Fig. 6 shows, the plot substantially deviated from the straight line; however, the limited number of experimental points did not allow consideration of this deviation as either systematic non-linearity or as random scatter. As a first approximation, we assessed this trend by linear regression. Being expressed as dependence on the echo attenuation, the trend and its standard error were −0.38 ± 0.17 dB/dB (r^2=0.71). Being expressed as dependence of the echo delay, the same trend was 3.46 ± 1.57 dB per delay doubling.

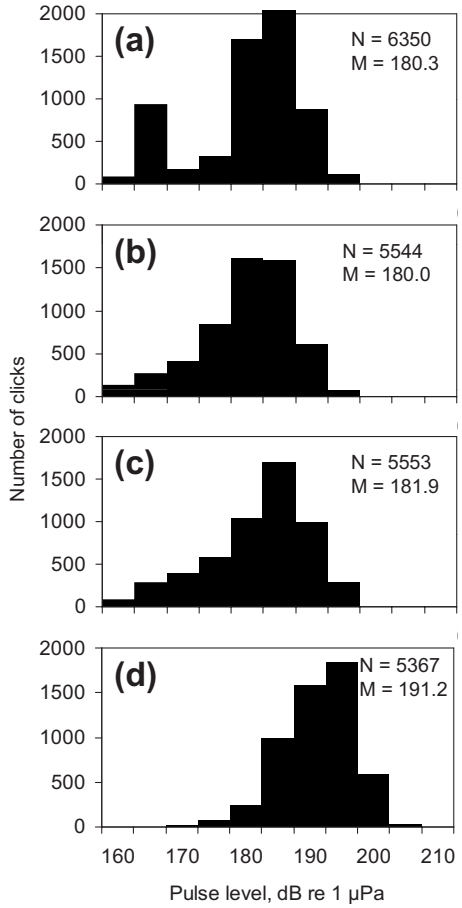


FIG. 5. Distributions of emitted pulse levels for four used combinations of echo attenuation and delay: (a) 2 ms, −31 dB; (b) 4 ms, −40 dB; (c) 8 ms, −49 dB; and (d) 16 ms, −58 dB. *N*—number of collected clicks and *M*—mean of the distribution.

Because of methodological restrictions (see Sec. II), it was important to find out how large the level difference was between successive sonar clicks. In order to do that, the level difference between each click and the next click was calculated. The result of the statistical analysis of 5274 click pairs from 300 randomly chosen click trains was conducted and is presented in Fig. 7. In some number of cases, the difference reached from −20 dB (a successive click was lower than the preceding one) to +20 dB (a successive click was higher than the preceding one). However, these sorts of cases were

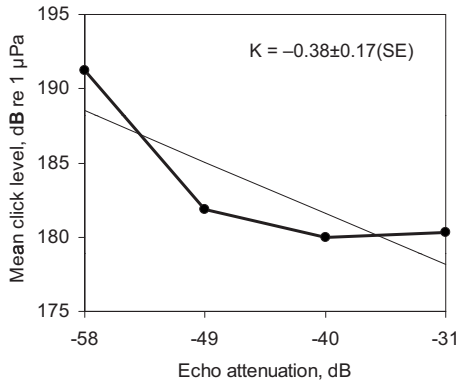


FIG. 6. Mean click level dependence on echo attenuation. Bold line—experimental data, thin straight line—regression, and *K*—regression slope.

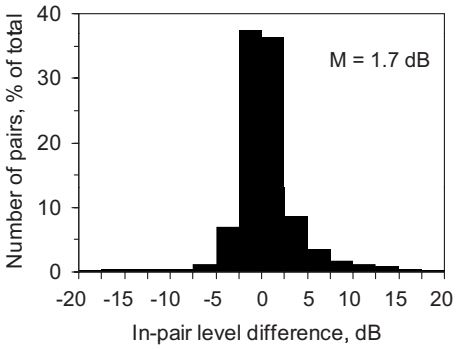


FIG. 7. Distribution of level difference between successive pulses. *M*—mean.

very rare. For the majority of the click pairs (73.8%), the difference between clicks was within a limit of ±2.5 dB, and for 89.3% the difference was within a limit of ±5 dB. This result shows that during each click train, the levels of clicks varied mostly in a gradual manner, with minor level differences between successive clicks.

C. AEP waveforms and magnitudes

All the obtained averaged AEP waveforms are presented in Fig. 8 at the four used combinations of echo parameters: (a) 2 ms delay and −31 dB attenuation, (b) 4 ms delay and −40 dB attenuation, (c) 8 ms delay and −49 dB attenuation, and (d) 16 ms delay and −58 dB attenuation. At each of the panels, AEP waveforms are presented at different levels of the emitted clicks, as obtained by sorting by click levels to 5-dB bins: from 170 ± 2.5 to 195 ± 2.5 dB re 1 μPa. Note that not all of the levels are presented in each panel. This is because not all bins contained enough original records (the number of available original records in each bin was equal to the number of collected clicks of corresponding level, see Fig. 5). If the bin contained less than 250 original records, the average record was omitted as not providing sufficient noise reduction.

All of the records presented here contained two characteristic AEP waveforms. The first one occupied a time window approximately lasting from the second to the fifth millisecond at the highest click level and shifting to the third to the sixth millisecond at the lowest click level, equally in all the records. This was interpreted as the response to the self-heard emitted biosonar pulse and further it is referred to as the emission-related response. The second waveform occupied a time window delayed relative to the first one by 2 ms (a) to 16 ms (d), according to the used echo delay. This was interpreted as the response to the echo and it is referred to as the echo-related response (Supin *et al.*, 2003).

D. AEP amplitude dependence on emitted click level and echo parameters

Peak-to-peak amplitudes of all emission-related and echo-related AEPs were measured. In doing this, we took into consideration that in spite of the large number of original records collected (in some bins, more than 1500, see Fig. 5), all the averaged records were substantially contaminated by background noise. The computation of rms values of presumably response-free pre- and post-response fragments of the records resulted in values of around 60–80 nV. Therefore, to specify the response amplitudes, a correction was made by subtracting the background noise power from the response power (see Sec. II). The peak-to-peak span was measured within 3-ms long windows; these windows are shown in Fig. 8. Note that the window offset increased with decreasing the click level with a rate of 0.15 ms per each 5-dB step (0.03 ms/dB). This was done because of an increase in AEP latency with decreasing stimulus level.

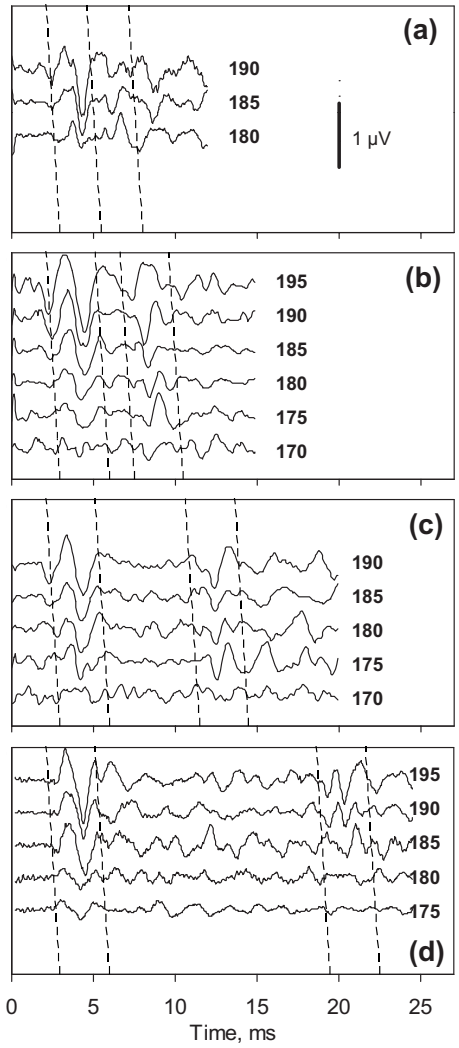


FIG. 8. Averaged AEP waveforms at four used combinations of echo attenuation and delay: (a) 2 ms, −31 dB; (b) 4 ms, −40 dB; (c) 8 ms, −49 dB; and (d) 16 ms, −58 dB. Emitted click levels for each of the records are indicated on the right (sorted in 5-dB steps with ±2.5 dB tolerance). Dashed straight lines delimit temporal windows where peak-to-peak span was assessed as an AEP amplitude. AEP amplitude calibration bar in (a) is common for all the records.

The results of measurements made in this way are presented in Figs. 9 and 10. The results revealed a substantial difference between the emission-related and echo-related responses.

For the emission-related AEPs, the only measurable variable influencing their amplitude was the level of the emitted click, so in Fig. 9 AEP amplitude is presented as a function of click level. The function demonstrated an obvious amplitude dependence on click level. This dependence

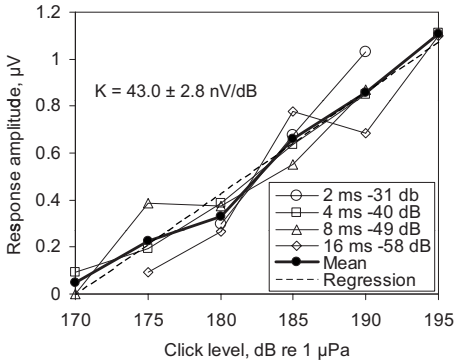


FIG. 9. Emission-related AEP amplitude dependence on emitted click level. Thin lines—plots for different combinations of echo attenuation and delay: (2 ms, −31 dB), (4 ms, −40 dB), (8 ms, −49 dB), and (16 ms, −58 dB), as specified in the legend; bold line—mean of the four plots, dashed straight line—regression, and *K*—regression slope.

was not much different between all four tested combinations of echo parameters; therefore it seemed reasonable to average these four plots. The averaged dependence could be successfully (r^2=0.98) approximated by a regression straight line with a slope of 43.0 ± 2.8 (SE) nV/dB.

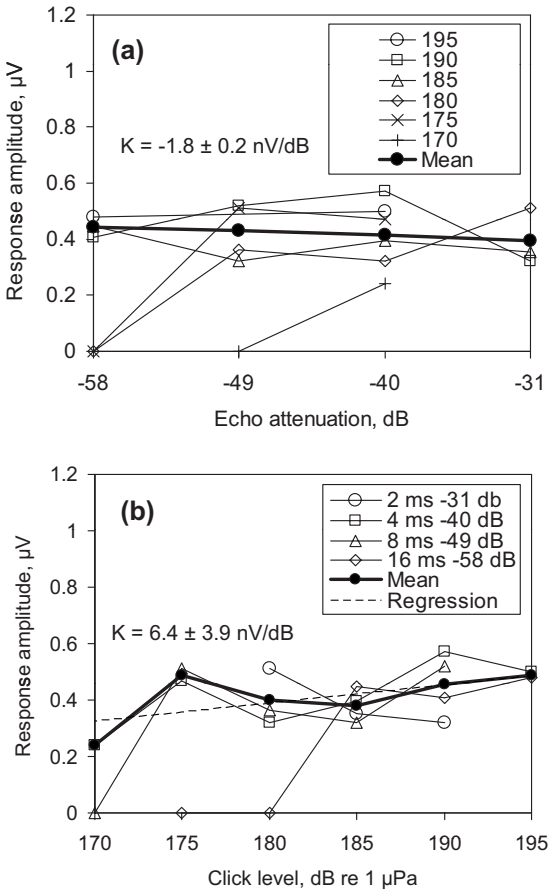


FIG. 10. Echo-related AEP amplitude dependence on echo attenuation and emitted click level. (a) Dependence on echo attenuation keeping click level as a parameter and (b) dependence on click level keeping echo attenuation as a parameter. Thin lines—plots for corresponding parameter values, as specified in the legends; bold line—mean of all the plots, excluding the points of AEP amplitudes below 0.2 μV; dashed straight line—regression; and *K*—regression slope. In (a) the regression line is not distinguishable from the mean plot.

The echo-related AEPs displayed quite different features. Since echo-related AEP depends on two variables, the echo attenuation and emitted click level, the results are presented in two forms: amplitude as a function echo attenuation, keeping the emitted click level as a parameter [Fig. 10(a)], and vice versa [Fig. 10(b)].

Functions presented in Fig. 10(a) demonstrated an absence of systematic amplitude dependence on echo attenuation within a wide range of variation of both echo attenuation and emitted click level. Except for three zero-amplitude points presenting combinations of the lowest values of both variables (−49 dB attenuation at 170 dB click level and −58 dB attenuation at 175 and 180 dB click levels), all of the other amplitude values were within a range from 0.24 to 0.57 μV and demonstrated random scatter rather than a systematic trend. Being averaged across emitted click levels (excluding the three zero-amplitude points), the amplitude dependence on echo attenuation was almost flat. It could be approximated by a regression straight line (in Fig. 10(a), it is not distinguishable from the mean plot) with a negligible slope of −1.8 ± 0.2 (SE) nV/dB.

Functions presented in Fig. 10(b) demonstrated almost no dependence of the echo-related AEP amplitude on the emitted click level within wide ranges of click levels and echo attenuations. Except for the three points mentioned above as combinations of the lowest values of the variables, AEP amplitude demonstrated more random variation rather than trend. Being averaged across different echo attenuations (again, excluding the three zero-amplitude points), the dependence on click level could be approximated by a straight regression line with a small slope of 6.4 ± 3.9 (SE) nV/dB.

IV. DISCUSSION

A. Validity of the echo attenuation control

As described in Sec. II, technical limitations did not allow us to compute and adjust attenuations of shortly delayed echoes based on the amplitude of the triggering emitted biosonar click. Instead, we controlled the echo attenuation based on the level of the click preceding the triggering one. We adopted this technique assuming that the click generating apparatus did not likely change the level very much from click to click when the clicks follow rapidly one after another. The statistical analysis of click levels confirmed this assumption. For the overwhelming majority of the biosonar clicks, the level difference between successive clicks did not exceed ±2.5 dB, and only in a few cases did this difference exceed ±5 dB (see Fig. 7). This means that for the majority of clicks, the error resulting from the substitution of the current click level by the previous click level could not exceed one bin-width of sorting click levels (±2.5 dB), and only in rare cases could the error be as large as one bin (±5 dB). In general, the systematic error should be even less because both positive and negative differences between successive clicks might compensate one another. In comparison with the overall range of click level variation (30–40 dB, see Fig. 5), we adopted the possible error as negligible as a first approximation.

B. AEP record quality and data scatter

In spite of large numbers of original records averaged (in some cases, more than 1500, in one case, more than 2000, see click statistics in Fig. 5), all the obtained averaged waveforms were substantially contaminated by background electrical noise (see Fig. 8). The noise frequency band overlapped the AEP band; so it could not be filtered out by regular frequency filtering. Therefore, the AEP records were of not the best quality.

Searching for a possible cause of the noise, it might be noted that AEP records obtained in the same subject and in very similar recording conditions (underwater voluntary positioning of the free moving animal with suction-cup electrodes) but in conditions of external sound stimulation (Supin *et al.*, 2007) were of much lower background noise, whereas AEP records obtained in the same subject in conditions of natural echolocation of real targets (Supin *et al.*, 2003, 2004, 2005) were noise-contaminated at approximately the same degree as in the present study. There might be a few reasons for the difference in noise level between the external-stimulation and echolocation conditions. First of all, the increased noise level might be a result of increased muscle activity during echolocation. This increased muscle activity might either result from the process of sonar click generation, or appear because the animal did not relax well enough during a short time (2–3 s) between the movement into the hoop station and movement back of the hoop to respond and signal the echo presence, whereas in external-stimulation experiments the animal stayed quietly relaxed in the hoop station for around 1 min of continuous brain-potential recording. The contribution of AEPs to multiple intra-head or external sound reflections should also be taken into consideration; however, no correlation of the noise level with the emitted pulse level was observed.

During the present investigation, we did not manage to reduce the background electrical noise down to the level characteristic of external-stimulation experiments. However, even in the background of the noise characteristic of the present study, definite AEPs were detectable and measurable, although the precision of measurements appears to have been reduced.

C. Echo-related AEP constancy: Implication to AGC problem

A remarkable feature of the data presented above is the nearly constant amplitude of echo-related AEPs within a wide range of echo attenuations and emitted click levels when the attenuation varied in association with echo delay. At a majority of the combinations of echo parameters and emitted click levels, the amplitude of the echo-related AEPs displayed random variation within a range from 0.24 to 0.57 μV but no systematic variation. Note that the echo *level* within these ranges of parameters varied as widely as by 38 dB: from 126 dB (175 dB emitted click level, −49 dB echo attenuation) to 164 dB (195 dB click level, −31 dB echo attenuation). At the same time, emission-related AEPs demonstrated obvious regular amplitude dependence on the stimulus levels within a markedly narrower level range of 25 dB (from 170 to 195 dB source level). Thus, the slight dependence of the echo-related amplitude on signal level is not an intrinsic feature of AEP; this is a feature of echo-related AEP in the current experimental conditions.

This specific feature of the echo-related AEP well confirms the hypothesis described in the Introduction. According to this hypothesis, the partial forward masking of an echo by the preceding emitted pulse may result in the constancy (or near-constancy) of a neuronal response to the echo, thus providing a kind of AGC within the auditory system of the echolocating subject. Indeed, the echo delay and attenuation in our experiments varied in a way that echo delay doubling was associated with 9-dB echo attenuation decrease. This is the same rate that characterized prolongation of forward masking with decreasing the test-to-conditioning click ratio in passive listening double-click experiments (Supin *et al.*, 2007) and echo detection threshold dependence on delay (Supin *et al.*, 2008). Exactly as predicted, the decreased echo attenuation, and release from forward masking due to increased echo delay, almost completely compensated one another when measured in an actively echolocating whale.

It is noteworthy that amplitudes of echo-related AEPs displayed an absence of dependence not only on echo attenuation but also on emitted click level. This effect can also be explained by features of the forward masking. Indeed, it has been shown in double-click experiments that duration of the forward masking depends on the stimulus level even if the conditioning-to-test ratio is kept constant. A proportional increase in both the conditioning and the test stimuli may result in a prolongation of forward masking; therefore, in spite of increasing the test stimulus level, the AEP amplitude may remain almost constant within a certain range (Supin and Popov, 1995). As these data indicate, the same effect took place during active echolocation by the animal. Increasing the emitted biosonar-click level should have resulted in a proportional increase in the echo level; however, within the range of action of forward masking, the amplitude of echo-related AEP remained nearly constant.

Of course, this mechanism of gain control is only capable of functioning while the echo level remains well above the absolute threshold of echo detection or a threshold of masking by the background noise. When the echo level is below the absolute or pseudo-absolute (in background noise) threshold, forward masking by the emitted click becomes useless, and it cannot function when there is nothing to mask. This can explain why combinations of the lowest values of echo attenuation and emitted click level in this experiment produced very low-amplitude AEPs. These combinations simply fell out of the AGC range. This may be made much clearer by presenting the echo-related AEP amplitudes as functions of not *echo attenuation* (as in Fig. 10) but absolute *echo level* (SPL) calculated in a simple way as

$$E = L - k,$$

where E is the echo SPL (dB re 1 μPa), L is the emitted click source level (dB re 1 μPa at 1 m), and k is the used echo attenuation factor (dB). Being presented in this way, the echo dependence appears as shown in Fig. 11. When the echo level was equal or above 125 dB re 1 μPa, the AEP

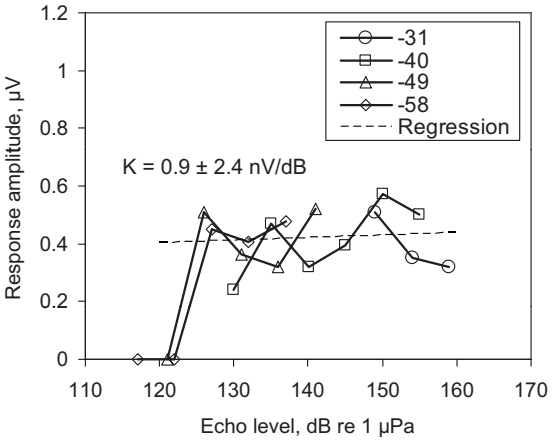


FIG. 11. Echo-related AEP amplitude dependence on echo SPL at four echo attenuation values, as specified in the legend. Dashed straight line—regression of all points except those below 0.2 μV, and *K*—regression slope.

amplitude fluctuated within a range 0.24–0.57 μV. The regression line drawn through these points had a negligible and statistically insignificant (less than the standard error) slope of 0.9 ± 2.4 (SE) nV/dB, i.e., it was almost flat. When the echo level approached a limit of 120 dB re 1 μPa, the AEP amplitude dropped down steeply. All of the three experimental points featuring zero AEP amplitude corresponded to the echo level nearly 120 dB re 1 μPa. This is the very same level that was determined as the AEP threshold for the subject whale during prior conditions of external stimulation (Supin *et al.*, 2007). So it is obvious that even a complete release from forward masking could not result in high AEP amplitude at echo levels so low. Thus, all the experimental data presented herein are in agreement with the hypothesis of the forward masking by emitted pulses as an AGC mechanism in the odontocete sonar.

It should be noted that in a study by Beedholm *et al.* (2006) AGC was not found in the auditory system of a harbor porpoise. However, AGC could not be observed with the experimental conditions that were used in that study. The echo-imitating signals were not proportional to the emitted biosonar signals but instead were kept at a relatively high level of 128 dB re 1 μPa and were triggered by biosonar pulses at a trigger threshold of 132 dB. This sort of relationship between the animal's emitted signal and the echo is not typical for the biosonar of odontocetes, which operates with echoes that are usually tens of decibels lower than the emitted biosonar signal. It has been shown (Supin *et al.*, 2006) that emitted biosonar pulses are heard by a whale at a sensation level down to −45 dB below the source level. Supposing, as a first approximation, that in the harbor porpoise the ratio is roughly the same, the emitted biosonar signals at the trigger level are perceived at the same sensation level as external signals at the level of 87 dB re 1 μPa. Even if the emitted pulses exceed the trigger level up to 40 dB, they still remain at a *lower* sensation level than the 128-dB electronic echo. In these sorts of conditions (where the conditioning signal is lower than the test signal), the forward masking would be negligible or absent. Since no forward masking would occur, no AGC would be expected. Thus, the results of Beedholm *et al.* (2006) do not contradict the hypothesis of forward-masking AGC but, to the contrary, actually confirm them. When the forward-masking mechanism was eliminated by an inadequate emission-to-echo ratio, AGC was, in fact, absent.

Therefore, we conclude that the forward masking of the echo by the preceding emitted pulse may well serve as an AGC mechanism for odontocete sonar.

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from the U.S. Office of Naval Research and the Russian Ministry of Science and Education. This is Contribution No. 1341 from the Hawaii Institute of Marine Biology. The whale work was completed under a Marine Mammal Permit issued to P.E.N. from the NMFS office of Protected Species with protocols approved by the University of Hawaii Animal Care and Utilization Committee. The valuable assistance of Dera Look, Stephanie Vlachos, and Vincent DePaolo is greatly appreciated.

Au, W. W. L., and Benoit-Bird, K. J. (**2003**). "Automatic gain control in the echolocation system of dolphins," Nature (London) **423**, 861–863.

Au, W. W. L., and Herzing, D. L. (**2003**). "Echolocation signals of wild Atlantic spotted dolphin (*Stenella frontalis*)," J. Acoust. Soc. Am. **113**, 598–604.

Au, W. W. L., and Würsig, B. (**2004**). "Echolocation signals of dusky dolphins (*Lagenorhynchus obscurus*) in Kaikoura, New Zealand," J. Acoust. Soc. Am. **115**, 2307–2313.

Aubauer, R., and Au, W. W. L. (**1998**). "Phantom echo generation: A new technique for investigating dolphin echolocation," J. Acoust. Soc. Am. **104**, 1165–1170.

Beedholm, K., Miller, L. A., and Blanchet, M. A. (**2006**). "Auditory brainstem response in a harbor porpoise show lack of automatic gain control for simulated echoes," J. Acoust. Soc. Am. **119**, EL41–EL46.

Bullock, T. H., and Ridgway, S. H. (**1972**). "Evoked potentials in the central auditory system of alert porpoises to their own and artificial sounds," J. Neurobiol. **3**, 79–99.

Nachtigall, P. E., and Supin, A. Ya. (**2008**). "A false killer whale adjusts its hearing when it echolocates," J. Exp. Biol. **211**, 1714–1718.

Popov, V. V., and Supin, A. Ya. (**1990**). "Auditory brain stem responses in characterization of dolphin hearing," J. Comp. Physiol. [A] **166**(3), 385–393.

Popov, V. V., Supin, A. Ya., and Klishin, V. O. (**2001**). "Auditory brainstem recovery in the dolphin as revealed by double sound pulses of different frequencies," J. Acoust. Soc. Am. **110**, 2227–2233.

Rasmussen, M. H., Miller, L. A., and Au, W. W. L. (**2002**). "Source levels of clicks from free-ranging white beaked dolphins (*Lagenorhinchus albirostris* Gray 1846) recorded in Icelandic waters," J. Acoust. Soc. Am. **111**, 1122–1125.

Supin, A. Ya., Nachtigall, P. E., Au, W. W. L., and Breese, M. (**2004**). "The interaction of outgoing echolocation pulses and echoes in the false killer whale's auditory system: Evoked-potential study," J. Acoust. Soc. Am. **115**, 3218–3225.

Supin, A. Ya., Nachtigall, P. E., Au, W. W. L., and Breese, M. (**2005**). "Invariance of evoked-potential echo-responses to target strength and distance in an echolocating false killer whale," J. Acoust. Soc. Am. **117**, 3928–3935.

Supin, A. Ya., Nachtigall, P. E., and Breese, M. (**2006**). "Source-to-sensation level ratio of transmitted biosonar pulses in an echolocating false killer whale," J. Acoust. Soc. Am. **120**, 518–526.

Supin, A. Ya., Nachtigall, P. E., and Breese, M. (**2007**). "Evoked-potential recovery during double click stimulation in a whale: A possibility of biosonar automatic gain control," J. Acoust. Soc. Am. **121**, 618–625.

Supin, A. Ya., Nachtigall, P. E., and Breese, M. (**2008**). "Forward masking as a mechanism of automatic gain control in odontocete biosonar: A psychophysical study," J. Acoust. Soc. Am. **124**, 648–656.
Supin, A. Ya., Nachtigall, P. E., Pawloski, J., and Au, W. W. L. (**2003**). "Evoked potential recording during echolocation in a false killer whale *Pseudorca crassidens*," J. Acoust. Soc. Am. **113**, 2408–2411.
Supin, A. Ya., and Popov, V. V. (**1995**). "Temporal resolution in the dolphin's auditory system revealed by double-click evoked potential study," J. Acoust. Soc. Am. **97**, 2586–2593.
Yuen, M. M. L., Nachtigall, P. E., Breese, M., and Supin, A. Ya. (**2005**). "Behavioral and auditory evoked potential audiograms of a false killer whale (*Pseudorca crassidens*)," J. Acoust. Soc. Am. **118**, 2688–2695.

Modeling cumulative sound exposure around marine seismic surveys

Christine Erbe[a)]
JASCO Research Ltd., 55 Fiddlewood Crescent, Bellbowrie, Queensland 4070, Australia

Andrew R. King
CSIRO-QCAT, Technology Court, Pinjarra Hills, Queensland 4069, Australia

(Received 11 September 2008; revised 27 January 2009; accepted 5 February 2009)

This article presents a method for reducing the computation time required for estimating cumulative sound exposure levels. Sound propagation has to be computed from every source position to every desired receiver location; so if there are many source positions, then the problem can quickly become computationally expensive. The authors' solution to this problem is to extract all possible source-receiver pathways and to cluster these with a self-organizing neural net. Sound propagation is modeled only for the cluster centroids and extrapolated for the entire geographic region. The tool is illustrated for the example of a marine seismic survey over a tropical coral reef. Resident fish species were expected not to flee the reef, but to stay among the corals for the entire duration of the survey. In such cases, the modeling of cumulative sound exposure levels is sometimes requested as part of environmental impact assessments. The tool developed combines a seismic source model, a near-field sound propagation model, and a far-field sound propagation model. The neural network reduces the computation time by a factor of 55. The cost is an error in modeled received levels of less than -1 ± 3 dB re 1 μPa^2 s.
 [DOI: 10.1121/1.3089588]

PACS number(s): 43.80.Nd, 43.30.Zk [WWA] Pages: 2443–2451

I. INTRODUCTION

To assess the likelihood of impact on marine mammals around seismic surveys, regulatory bodies generally require the modeling of sound exposure levels (SELs) around single shots.[1–3] From pre-defined exposure thresholds, maximum ranges are determined, over which visual observation and a number of mitigation methods are suggested. With common techniques such as ramping-up and shutting-down if marine mammals are detected within a proximal zone, it is hoped that marine mammals will not be exposed to harmful sound levels. One underlying assumption is that marine mammals will try to avoid zones of potentially harmful levels, if possible. If the area is determined to constitute critical habitat though, animals might not flee and a desertion of habitat might not be "desired," therefore regulations differ.[1]

In the case of fish, some studies have reported a desertion of the area ensonified by seismic sounds,[4–6] measured as a decline in catch rate; occasionally the abundance of fish increased tens of kilometers away from the source. Slotte *et al.*[7] showed that fish descended to deeper water when airguns operated above. Some fish species are "attached" to a particular habitat and do not suddenly flee tens of kilometers in the horizontal to escape sound exposure. This can be the case in river deltas[8] or coral reef environments.[9] Both environments have shallow water. In the case of tropical coral reefs, the water depth is usually less than 25 m.[10,11] Reef fish that simply hide among coral when "danger" is perceived from above will not be able to escape sound exposure from a seismic survey. For this reason, environmental impact assessments in such areas are beginning to look at cumulative effects from multiple exposures.

Within the bioacoustic research community, cumulative effects are still poorly understood.[12] Hastings and Popper[13] summarized the published effects on fish (behavioral responses, physiological responses, hearing loss, and tissue damage) until 2005. Yelverton *et al.*[14] exposed fish to blasts, identifying a relationship between the total energy in the blast and mortality due to organ damage. Stuhmiller *et al.*[15] summarized data on blast injury to the lung in terrestrial animals of different sizes and found that the incidence of injury and lethality correlated with total energy. McCauley *et al.*[16] looked at effects from one seismic airgun towed toward and away from caged fish. Fish ears experienced substantial damage after cumulative exposure. Govoni *et al.*[17] concluded that acoustic trauma in fish was proportional to total energy in acoustic impulses. Popper *et al.*[8] exposed fish to 5 airgun shots and 20 shots, showing a higher temporary threshold shift after 20 shots, hinting at a cumulative effect.

Estimating sound exposure requires the modeling of sound propagation through the marine environment. Etter[18] tabulated a range of sound propagation models as they existed in 2003. Sound propagation models require certain environmental input parameters such as sound speed profiles for the water column, sound speed and attenuation data for the seafloor, and bottom topography measurements. Data on seafloor geoacoustics are sometimes hard to come by, but in many environments, the geology does not vary drastically over the required modeling ranges and one profile (e.g., obtained from bore hole measurements prior to full-scale seis-

a)Electronic mail: christine@jasco.com

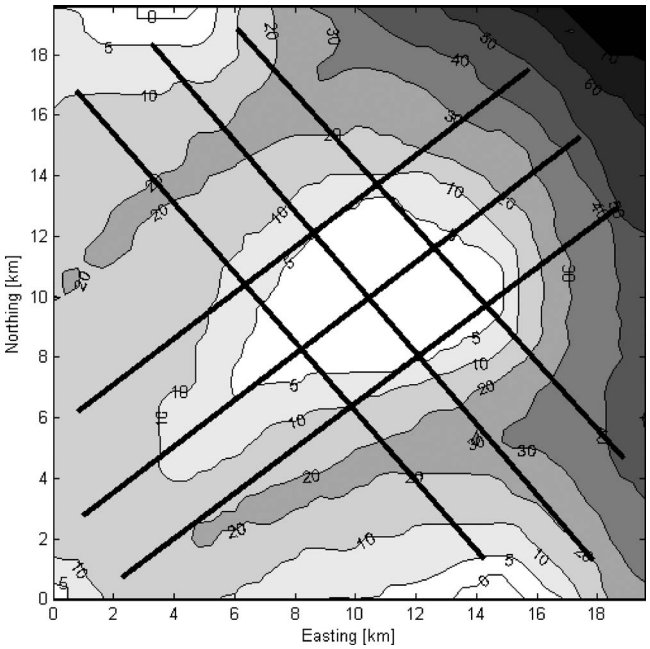


FIG. 1. Island reef system with six seismic transects. Depths in meters.

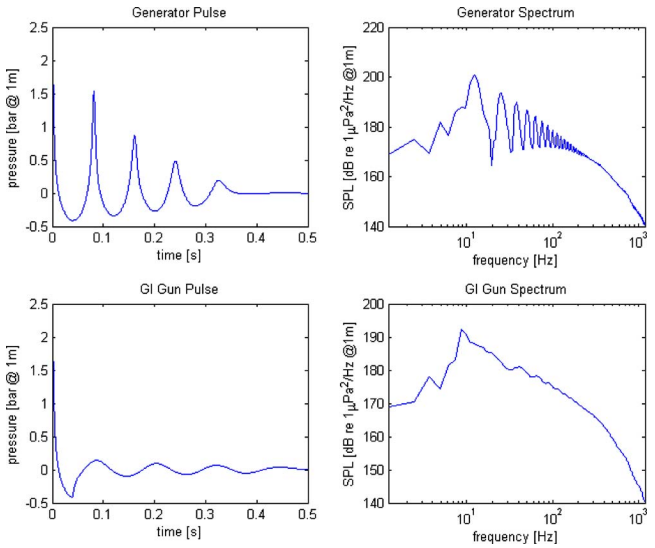


FIG. 2. (Color online) Top: Waveform and power density spectrum of one generator chamber (45 in.^{3}). Bottom: Waveform and power density spectrum of one GI gun (150 in.^{3}) including both the generator (45 in.^{3}) and the injector (105 in.^{3}) fired 42 ms after the generator.

mic surveys) suffices. Sound speed profiles for the water column are usually easier to obtain; they vary with time of day and season and should be matched to the planned operation. Many sound propagation models are range-dependent and able to model changing acoustic parameters with range. Bathymetry is often known in quite some detail and can vary significantly in all directions throughout an operational area. Particularly in downward refracting (summer) water, bathymetry can play a major role in sound propagation. Modeling sound propagation from all shots in a survey to all possible receiver locations is computationally very demanding. Sound propagation has previously been modeled for a small number of "representative" pathways (bathymetries) and has been applied to the entire survey.[19–21]

This article presents a method of finding an optimum set of representative bathymetries. The method can be used for cumulative sound exposure modeling in environments that are homogeneous with regard to sound speed profile and seafloor geology. The algorithm to estimate cumulative SELs from an entire seismic survey is detailed. This article does not attempt to assess biological impact on marine fauna from cumulative exposure.

II. THE STUDY SITE

The study site was an island reef system (Fig. 1). The mainland lay to the south-west. The continental shelf dropped rapidly toward the north-east. In the center of the study site lay a coral reef that was about 5 m below the water surface at high tide when the survey was planned. Coral was growing on limestone bedrock. The proposed marine seismic survey encompassed six transects. The combined transect length was 121 km. The shot spacing was 25 m, yielding a total number of 4855 shots. The intended shot timing was about 10 s (speed of the towing vessel: 5 kn).

III. MODELING THE SOURCE

The seismic array consisted of ten identical generator-injector (GI) guns arranged in two lines. Each line had five guns separated by 3 m. The spacing between the lines was 5 m. A sketch of the array is shown in Fig. 8, Sec. VI. The tow depth was 2 m. To model the acoustic pressure, we used a full-waveform airgun array source signature model,[22] which was based on the physics of the oscillation and radiation of airgun bubbles.[23] The model solved a set of parallel differential equations, accounting for non-linear pressure interactions between airguns, port throttling, bubble damping, and GI-gun behavior.[24–26] GI guns have two chambers, a generator chamber, which is fired first, and an injector chamber, which is fired at half the generator bubble period (42 ms), in order to reduce the unwanted bubble pulse (Fig. 2). The frequency spectrum of a GI gun is smoothed.

All ten airguns were modeled simultaneously, while accounting for pressure interactions between neighboring guns. The resulting radiation directivity pattern of the array is shown for the horizontal plane in Fig. 3. The source level of the array is a function of azimuth, elevation, and frequency. We computed levels for 24 adjacent 1/3 octave bands between 10 Hz and 2 kHz. The outer thick black line corresponds to the broadband directivity pattern. Integrated over all frequencies, the array resembled a point source (monopole) in the horizontal plane with a source level of 217 dB re 1 μPa. Moving further toward the inside of the polar plot in Fig. 3, the radiation pattern is shown for selected 1/3 octave band center frequencies. At low frequencies, the array behaved like an acoustic monopole source. At about 80 Hz, the radiation pattern became dipolar, at 120 Hz quadrupolar, and for higher frequencies multipolar.

IV. MODELING SOUND PROPAGATION

The geoacoustic properties of the seafloor were taken from standard reference works and research articles[27–32] as

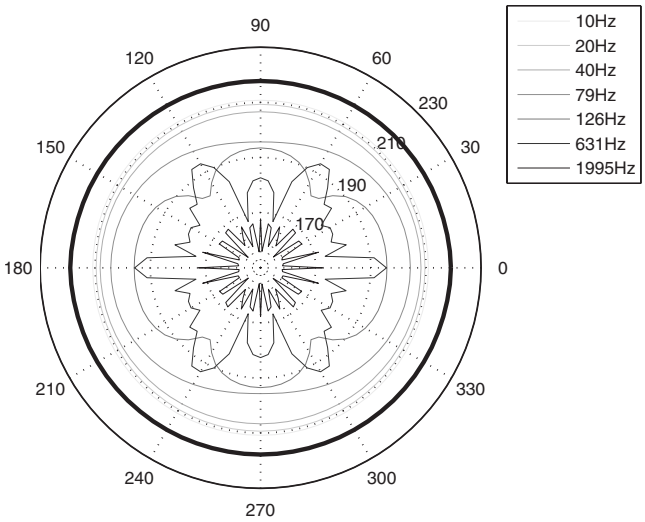


FIG. 3. Directivity pattern of the seismic array in the horizontal plane. Plotted are source levels (decibels re 1 μPa) for seven selected $\frac{1}{3}$-octave bands as a function of azimuth. The tow-direction is along the positive x-axis (0 deg). The outermost thick black line with a radius of 217 dB represents the broadband source level.

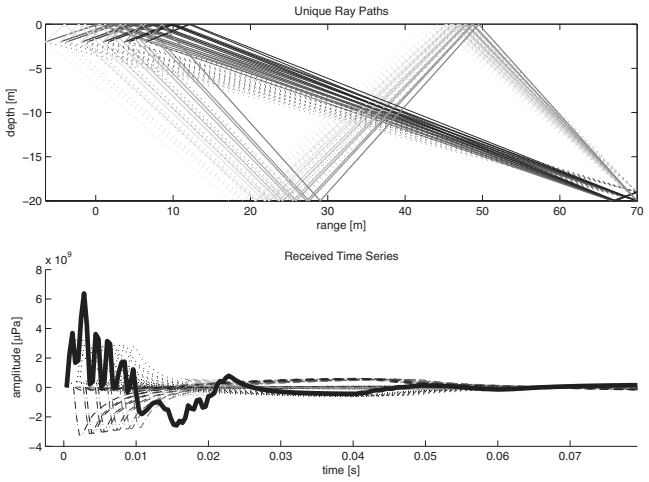


FIG. 4. Eigenray paths from the ten airguns at 2 m depth to a receiver at 19–20 m depth and 70 m range. There are 80 rays, 8 per airgun: (1) downward-direct, (2) downward—1 bottom reflection, (3) dwd.—1 bot. refl.—1 surface refl., (4) dwd—1 bot. refl.—1 sfc. refl.–1 bot. refl., (5) upward—1 sfc. refl., (6) upwd.—1 sfc. refl.—1 bot. refl., (7) upwd.—1 sfc. refl.—1 bot. refl.—1 sfc. refl., and (8) upwd.—1 sfc. refl.—1 bot. refl.—1 sfc. refl.—1 bot. refl. The black line in the bottom picture shows the total received pulse as a sum over all 80 eigenray pulses.

well as bore hole measurements in similar environments. For the example in this article, an averaged and simplified bottom was modeled with constant S and P wave attenuations of 0.3 and 0.2 dB/λ, respectively. Shear speed was modeled as 500 m/s at the seafloor; compressional wave speed increased from 2000 m/s at the seafloor surface to 4000 m/s at 2000 m below the seafloor surface. Density increased linearly from 2.0 to 2.6 g/cm^3 over the first 2000 m of seafloor. Sound speed profiles of the water column were obtained from a local database.

A. In the near-field

Over the first 100 m around the airgun array, we used the ray tracing model RAY (Ref. 33) with added functionality for bottom and surface reflection and scattering losses, and absorption (molecular relaxation).[34] RAY shoots a fan of rays from the source and propagates all rays out to a desired range and depth by repeatedly applying Snell's law. RAY was run from each airgun in the array to a receiver grid in space. Eigenrays were searched for; they have a unique path characterized by the source-departure-angle (up or down) and the number of top and bottom turns and reflections. Turns are due to refraction in a sound velocity gradient. Received sound levels (or transmission losses) were computed as the sum over all eigenrays for all gun-receiver pairs.

RAY treated the seafloor as an infinite layer, at the boundary of which acoustic energy was lost. In many marine environments, the upper seafloor consists of unconsolidated, water-saturated sediment. Loss is calculated as a two-fluid-layer model.[34] At the coral reef, however, the seafloor showed limestone bedrock. The bottom boundary was therefore modeled as a fluid-to-solid boundary, using Eq. (4.25) in Ref. 35 for the reflection coefficient.

For sound sources near the water surface (such as an airgun array towed at a few meters depth), a phenomenon called the surface ghost is important. Sound emitted upward reflects on the water-air interface and overlaps with sound emitted downward. Depending on departure angles, the signals cancel out at certain depth-range combinations in the near-field, and they interfere constructively at others. We accounted for a π phase shift with each surface reflection.

The code took the pressure waveform of each airgun (Fig. 2) and computed the Fourier transform. From this, the transmission loss (TL) was subtracted in the frequency domain at each receiver grid point for each airgun source and for each eigenray. The inverse Fourier transforms were computed, and all eigenray pulses were superposed in the time domain, after applying a time shift depending on the ray travel time (Fig. 4). This yielded the overall received pulse as a function of time at each receiver grid point. The SEL corresponds to the total energy of a signal. It was computed from the pressure pulse time series and measured in decibels re 1 μPa^2 s:

$$\mathrm{SEL} = 10\log_{10}\left(\int_T p(t)^2 dt/p_{\mathrm{ref}}^2\right).$$

B. In the far-field

To model sound exposure at long ranges, we used the Range-dependent Acoustic Model (RAM) (Ref. 36) for the current example. The code solved the parabolic equation using a split-step Padé algorithm[37] and Collins' parabolic equation (PE) self-starter.[38] Energy loss into shear wave propagation was accounted for by applying a complex density approach.[39] The horizontal radiation pattern (Fig. 3) was used for the frequency- and azimuth-dependent source levels in long-range propagation. Figure 5 shows an example of a RAM output: received levels as a function of range and depth. The left graph shows the sound speed profile for the water, which was used for the entire reef. Where the seafloor

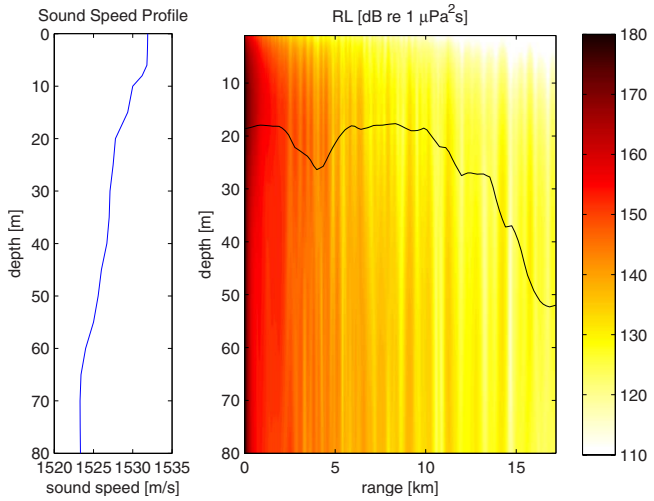


FIG. 5. (Color online) TL as a function of range and depth from one shot of the array. The left graph gives the sound speed profile in the water down to the maximum depth of 80 m in this reef environment.

was shallower than 80 m, the water profile was truncated and replaced with the *P*-velocity profile. The black line in the right plot shows the bathymetry for the example run.

V. MODELING CUMULATIVE SEL: FAR-FIELD

A. Computing a library of bathymetries

1. Generating a subset of shot-receiver pairs

A subset of shot-receiver pairs was generated by sampling shots at 800 m intervals (every 32nd shot), resulting in 154 sampled shot points. A coarse receiver grid was generated, covering the area, with a spacing of 1600 m in both directions, easting and northing. This resulted in 166 receiver grid points.

All receiver points were matched with all shot points, yielding 25 564 shot-receiver-pairs. The distances between all shots and all receivers were computed. All shot-receiver-pairs, which were more than 15 km apart, were discarded under the assumption that shots that far away would not contribute significantly to the cumulative SEL at the particular receiver location. Furthermore, all pairs with distances shorter than 1 km were discarded as these were too short to be useful in a "library" of representative shot-receiver propagation paths. It is reasonable to assume that the bathymetries associated with shot-receiver distances of less than 1 km are automatically included in the longer shot-receiver distances.

At this stage, the number of shot-receiver-pairs was 22 084. For each of these pairs, a bathymetry profile was generated from Fig. 1, by spline-interpolation, to an along-profile sampling distance of 250 m. All profiles were smoothed by applying a convolutional filter with a rectangular window. The spatial window was shorter at the beginning and end of each profile, in order to hold the depths of the start- and end-points fixed. Otherwise, effective zero-padding would bring the end-points closer to the surface than they are in reality.

These bathymetry profiles were considered a representative base of all shot-receiver profiles in the survey, from which an optimized library could be computed.

2. Cluster-analysis of bathymetries

A self-organizing map (SOM),[40,41] also called an unsupervised Kohonen neural network, was used to cluster the 22 084 bathymetry profiles into similar groups. The SOM consisted of 30×30, i.e., 900 neurons, arranged in a hexagonal topology with toroidal shape (wrapped top-to-bottom and side-to-side). After training, each neuron or cell in the SOM "contained" a cluster of similarly-shaped bathymetries, represented by a median or central profile.

To further reduce the number of representative profiles, we utilized *K*-means clustering.[42] It is a fast method of assigning vectors to cluster centers (called centroids) by comparing Euclidean distances. The 900 central SOM profiles were thus clustered into 64 clusters. The centroids were computed as the arithmetic mean of each cluster. Figure 6 shows all 900 SOM profiles (gray) grouped into 64 clusters, and the centroid bathymetries (black) for each cluster. These centroids made up the final library of bathymetries for which sound propagation was modeled.

B. Modeling TL along the library bathymetries

The far-field sound propagation model was run for each of the 64 centroid bathymetries. This yielded 64 matrices of TL as a function of range and depth for each of the 24 frequencies modeled.

C. Assigning library TL matrices to all shot-receiver-pairs

For the computation of the library bathymetries, a shot sampling of 800 m (every 32nd shot) and a receiver grid of 1600 m resolution in both directions (easting and northing) had been used. For the actual computation of cumulative sound exposure, a shot sampling of 400 m (every 16th shot) and a receiver grid spacing of 250 m were used. All shots were matched with all receivers and bathymetry profiles extracted from Fig. 1. Each extracted bathymetry was passed through the Kohonen classifier and *K*-means algorithm to assign the best bathymetry centroid for which TL had been computed.

D. Integrating over all shots

Knowing the array heading at each shot, the source level of the array in the direction of the current shot-receiver-pair was taken from the directivity patterns (Fig. 3) for each of the 24 frequencies modeled. The TL for each frequency was subtracted (in logarithmic terms, i.e., divided in linear terms) along the shot-receiver path. Energy was integrated over all frequencies to yield a broadband SEL at all depths at the current receiver range. This was done for all shot-receiver-pairs.

Finally, at the receiver grid, energy was integrated over all modeled shots. Only every 16th shot had been modeled. Accounting for the missed shots corresponded to four doublings in energy, or an addition of 12 dB to the computed SEL. The resulting sound exposure is illustrated in Fig. 7. The levels were plotted near the seafloor (motivated by the assumption that fish would hide among the coral). Horizontal

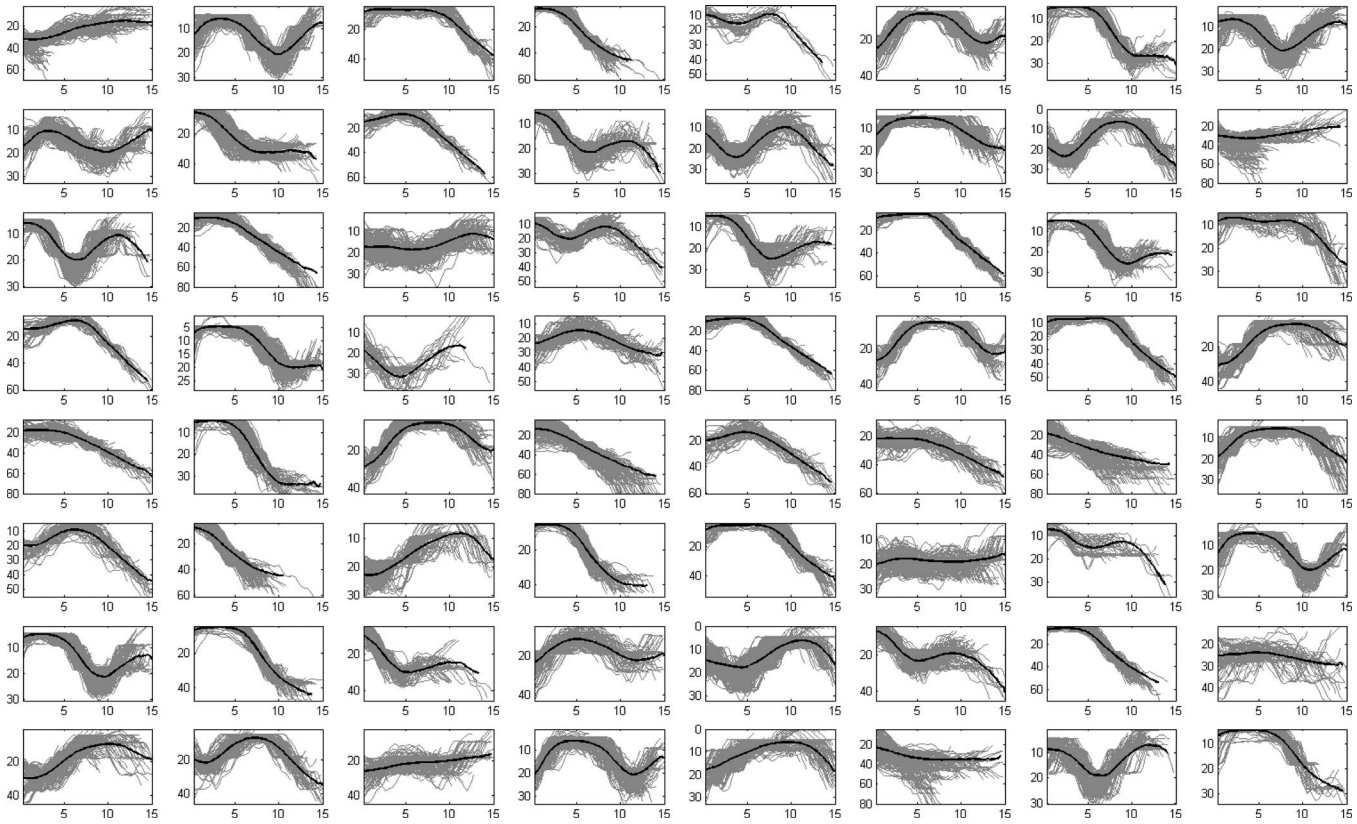


FIG. 6. 64 clusters of bathymetries showing the centroid profile in black and the 900 SOM profiles in gray. *Y*-axes: depth (in meters); *x*-axes: range (in kilometers).

sound propagation in shallow water is very poor. Therefore, SELs drop fast in shallow water. SELs were low over the reef in the center of the figure and toward the islands in the north-west and south-east.

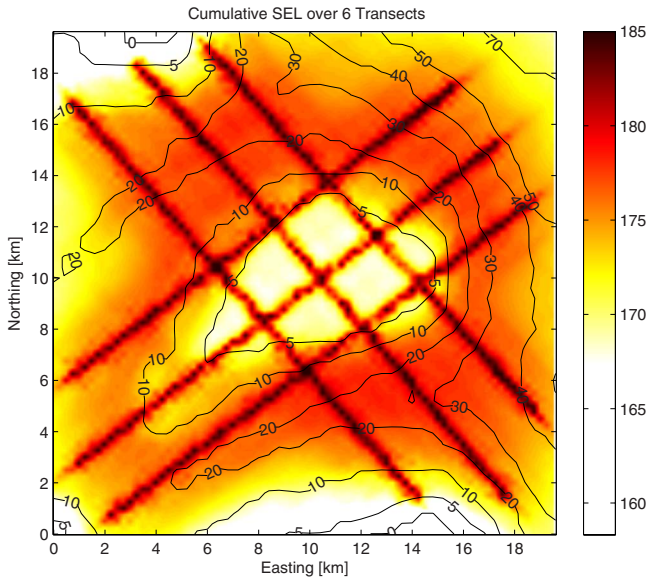


FIG. 7. (Color online) Cumulative SEL (decibels re 1 μPa^2 s) from the entire survey plotted at the seafloor. Given that the far-field propagation model was used, levels directly underneath the transect lines should be considered inaccurate; the reader should instead refer to the near-field results in Fig. 8.

VI. MODELING CUMULATIVE SEL: NEAR-FIELD

A. Near-field to far-field transition

The far-field model calculated TL from an equivalent point-source to receiver locations at various distances, depths, and bearings. However, a seismic array consists of many sources and so the point-source assumption is not valid in the near-field. The maximum extent of the near-field of an array can be estimated by the expression[43]

$$R_{\mathrm{nf}} < \frac{L^2}{4\lambda},$$

where λ is the acoustic wavelength and L is the longest dimension of the array. Along the diagonal of the modeled array, $L \approx 13$ m, and so the maximum near-field range is 28 m at 1 kHz (R_{nf} is less for lower frequencies).

B. Acoustic imprint of the array on the seafloor

RAY was run with a flat seafloor at a series of depths. The source depth (airgun array location) was 2 m below the water surface. Results are shown for four depths in Fig. 8. Plotted were SELs from one shot in the center of the picture as received at the seafloor. The ten airguns in the array were indicated as circles. Contours were drawn at SELs of 180, 185, 190, 195, and 200 dB re 1 μPa^2 s, if these levels were reached.

At seafloor depths of 5 and 10 m, SELs were in excess of 200 dB re 1 μPa^2 s underneath the array. At 20 m depth,

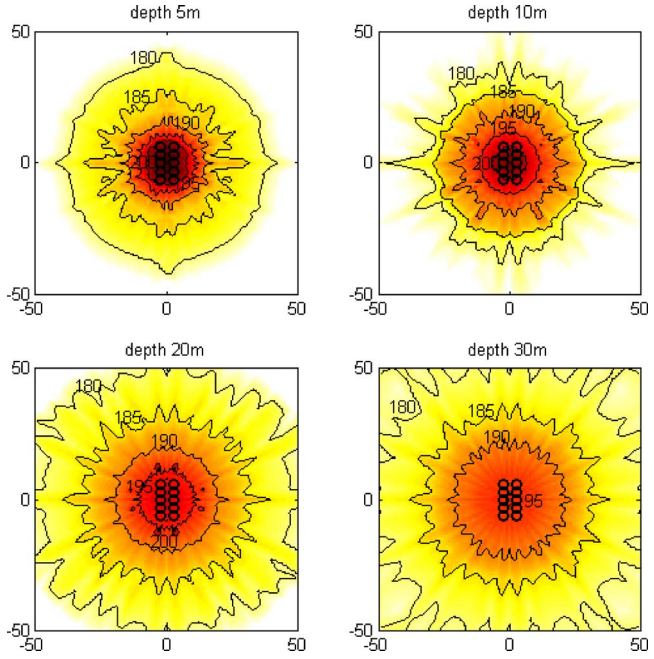


FIG. 8. (Color online) Acoustic imprint of the array at the seafloor. The tow-direction would be straight up, due north in this figure. The ten airguns of the array are shown as circles in the center. Received SELs at the seafloor are plotted for four different seafloor depths. Contours are drawn at 180, 185, 190, 195, and 200 dB re 1 μPa^2 s. *X*- and *y*-axes in meters.

200 dB re 1 μPa^2 s was only reached at four points slightly ahead and behind the array; the tow-direction was upward, or due north. At 30 m depth, levels were below 200 dB; 195 dB re 1 μPa^2 s was reached in a small region directly underneath the center of the array. High SEL disappeared as the seafloor got deeper. In these tropical waters, the sound speed profile was downward refracting. As a result, the decrease in SEL with range was faster over shallow depths. Therefore, the 180 dB contour was closer to the source at 5 m depth than at 30 m depth.

C. Integrating over one transect

Figure 9 shows SEL in shallow water integrated over short pieces of seismic transect. Levels were plotted at the seafloor, which ranged in depth from 5 to 30 m. At 5 and 10 m depths, levels were in excess of 200 dB re 1 μPa^2 s only directly underneath the array. The 195 dB contours (and lower) of two separate shots joined to form bands running parallel to the transect line. At 20 m depth, 200 dB was reached only at two small areas in between two subsequent shots. At 30 m depth, levels were less than 200 dB.

VII. ENSONIFIED AREAS

The model can be used to estimate how much of a reef (what percentage of area) receives certain SELs. One way of doing this is to augment the far-field SEL data with the near-field SEL data and to replot Fig. 7 at a finer resolution. The resulting map can then be overlaid with Geographic Information System (GIS) based habitat maps to estimate ensonified areas. Most of the time, the concern is about high SELs, which are only reached close to the source. Only the near-field results (Fig. 9) are relevant then for the computation of

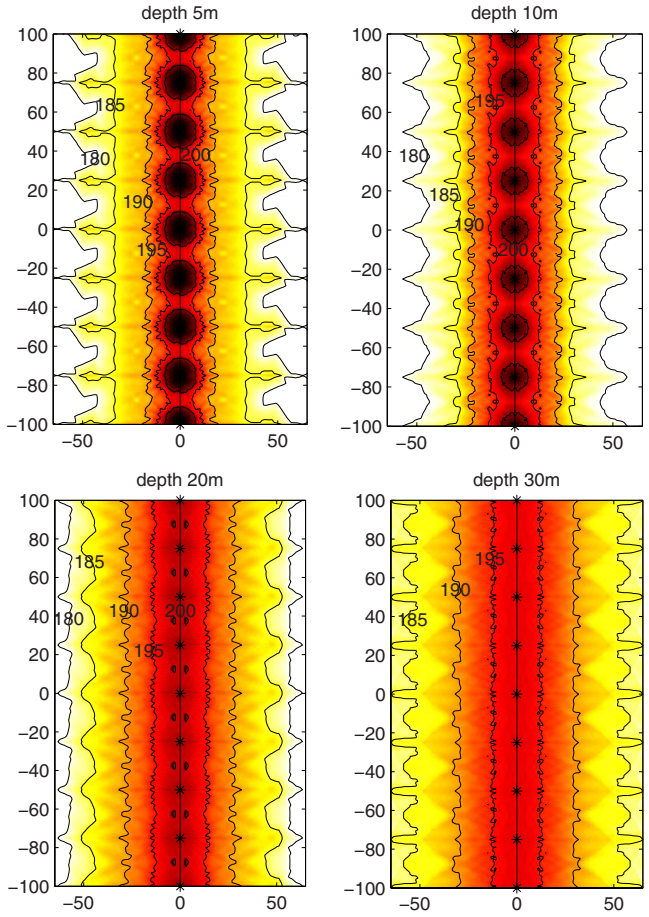


FIG. 9. (Color online) Cumulative SEL underneath short pieces of seismic transect over different seafloor depths. Contours are plotted for 180, 185, 190, 195, and 200 dB SELs, where these levels were reached. Each picture contains nine shots (*) along one transect. *X*- and *y*-axes in meters.

ensonified areas. The model computed the mean area per shot from Fig. 9 for any SEL threshold. In Table I, the areas per shot that received SEL in excess of 190, 195, and 200 dB re 1 μPa^2 s are listed for six different seafloor depths. As the seafloor got deeper, high SEL contours disappeared. Due to the downward refracting acoustic environment, the area of the 195 dB contour increased before it decreased to disappear. The same behavior would be found for the 190 dB contour if the survey was carried out over deeper seafloor depths.

Multiplying the area per shot by the number of shots occurring over different seafloor depths yielded the total ensonified area for each of the three SEL thresholds. Defining the boundary of the reef as the outermost 50 m depth isopleth, the total reef area in Fig. 1 was 376 km^2. The areas within each SEL contour are summarized in Table II.

VIII. COMPUTATIONAL COSTS AND BENEFITS

The cumulative SEL model was programed in MATLAB (The MathWorks Inc.) on a dual core, 2.5 GHz, laptop with 4 Gbyte RAM, running Windows Vista (Microsoft Corp.). It took 31 h from the extraction of shot-receiver-pairs, via the neural net clustering and propagation modeling, to the computation of Fig. 7. By comparison, using the far-field model to compute SEL over the entire reef from only one shot took

TABLE I. Cumulative SEL over different seafloor depths.

Seafloor depth (m)	Area per shot within 190 dB contour (m^2)	Area per shot within 195 dB contour (m^2)	Area per shot within 200 dB contour (m^2)
5	823	460	223
10	1100	568	136
20	1390	674	12
30	1518	571	0
40	1713	425	0
50	1760	0	0

21 min. Modeling all 4855 shots would take 1700 h =71 days. The neural-network-based algorithm is thus a factor of $1700/31 \approx 55$ faster. The increase in speed comes at a cost in accuracy.

In order to assess the accuracy of the SOM-based model, both RAM and the SOM-based algorithm were run for a single shot indicated by an asterisk in Fig. 10. The resulting SELs of the SOM were subtracted from those of RAM. For most receivers, the difference was less than 1 dB re 1 μPa^2 s. The SOM performed well if the bathymetry did not change rapidly; deviations were largest on the slopes of the reef.

IX. DISCUSSION

This article presented a technique to map cumulative sound exposure over a large geographic area from a full marine seismic survey. Modeling each shot of a survey individually is computationally very demanding. Instead, a method was presented to subsample all shot-receiver-pairs and to compute an optimal library of representative shot-receiver paths. This was done by grouping all shot-receiver paths with a SOM (Kohonen neural network) and subsequent *K*-means clustering. Widely-used sound propagation models were used to compute sound propagation along the cluster centroids: ray tracing in the near-field and a parabolic equation model in the far-field. Propagation results for the centroids were then matched with all shot-receiver-pairs, and energy was integrated over all shots on a grid of receiver locations spanning the geographic area. Mapping cumulative SEL with the neural-net-based algorithm reduced the computation time significantly. This came at a cost of error in SEL of <1 dB re 1 μPa^2 s, on average, with a standard deviation of <3 dB. The algorithm worked well in areas where the bathymetry did not change rapidly.

In a previous study of ours, cumulative SEL was modeled for a seismic survey over a coral atoll, which was flat and shallow on the inside, but had steep slopes on the outside. The SOM did not manage to cluster all of the bathymetries satisfactorily at once. Two separate SOMs were used, one for flat and shallow bathymetries, the other for steep and deep bathymetries, significantly improving the classification. More than one SOM will also need to be run if the area is not homogeneous with regard to water sound speed and geoacoustics. If the sound speed profile or the geology varies significantly over the survey area, the geographic region will need to be split into a set of homogeneous environmental provinces, and a SOM will need to be run for each province.

If sound speed profiles in the bottom or water are not known or are known to vary significantly, it is good practice to run the model with the extremes of the profiles to compute upper and lower bounds of expected sound exposure.

The focus of this study was to develop an efficient method to compute cumulative SEL over large areas and not to show absolute levels from some survey over some reef. The examples presented were run with an averaged and sim-

TABLE II. Ensonified reef areas.

Category	Area receiving this level (km^2)	% of total reef area
SEL$>$190 dB re 1 μPa^2 s	5.80	1.54
SEL$>$195 dB re 1 μPa^2 s	2.72	0.72
SEL$>$200 dB re 1 μPa^2 s	0.48	0.13

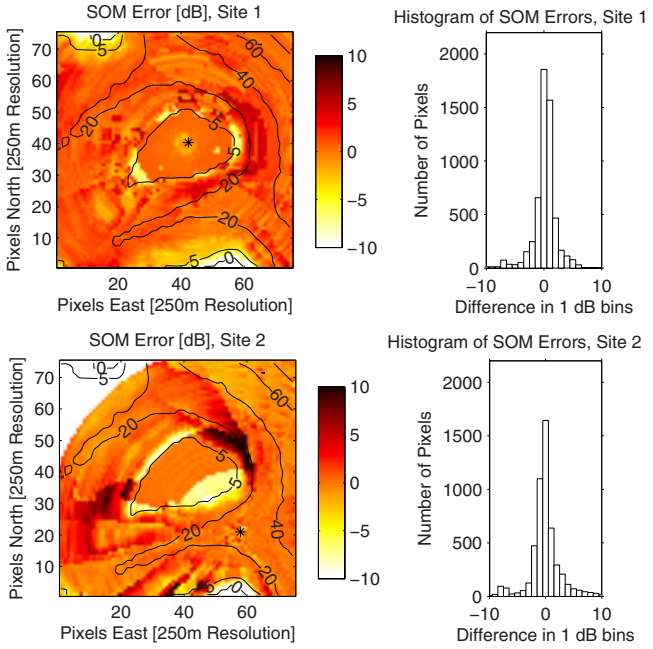


FIG. 10. (Color online) Error analysis of the neural-net-based algorithm. Shown is the difference in modeled SEL between a three dimensional run of RAM and the SOM. Results are shown for two sites; the asterisk indicates the source location. From the histogram, the difference was -0.5 ± 2.0 dB re 1 μPa^2 s for a source in the center of the reef (site 1) and -0.1 ± 2.7 dB re 1 μPa^2 s for a source south-east of the reef (site 2). The white segment in the top-left of the plot at site 2 was beyond the maximum modeled range of 15 km.

plified bottom. If full seafloor velocity profiles are available from bore hole data, then these should be used instead. We ran RAM in our example because it is beautifully stable and because the wave speeds were low enough near the top of the seafloor to apply the complex density approximation[39] for shear wave loss. For limestones of higher sound speeds and lower attenuations, and if modeling to very low SEL (long ranges) is desired, a model that also propagates *S*-waves through the bottom [e.g., RAMS (Ref. 44)] should be used. An acousto-elastic model might also be preferable for coral reefs with large outcrops that can reach the surface and strip away waterborne energy.

The algorithm presented has repeatedly been used to estimate ensonified areas in situations where environmental agencies asked what percentage of a reef received potentially damaging levels. It has to be pointed out that the use of the SEL metric for this generates a single exposure "equivalent" assuming no recovery in between exposures, likely overestimating effects. While any threshold can be modeled, we continuously point out that the bioacoustic impact of cumulative sound exposure is still poorly understood.

Our algorithm computed SEL as the integral of squared pressure over time. Strictly speaking, this relationship is only valid in the far-field. The total energy of a pulse is the integral of the intensity over the duration of the pulse and over the (spherical) surface that it passes through. Acoustic energy consists of kinetic energy and potential energy. The former is contained in particle movement; the latter is equal to the work done by elastic pressure forces. Intensity is the product of pressure and particle velocity. The pressure gradient is proportional to the particle acceleration[45]

$$\nabla p = -\rho \frac{\partial u}{\partial t}.$$

In the far-field, $p = \rho c u$, with p denoting pressure, ρ density, c speed of sound, and u particle velocity. Putting it all together, in the far-field, the energy per unit area (also called energy flux) becomes

$$E = \frac{1}{\rho c} \int p^2 dt.$$

Given that the acoustic impedance ρc can mostly be considered constant in the water column, the first term is usually ignored.

In the near-field, particle velocity has a "quadrature" component that is out-of-phase with acoustic pressure.[45] When talking about acoustic energy, particle velocity and pressure should correctly be both included in the near-field.

What actually are the "damaging" sound characteristics is still not fully understood. Likely, the characteristics will differ depending on what type of "damage" is considered. Is it sound pressure level above sensitivity, as in the linear threshold shift (LINTS) hypothesis?[8,46] Is it intensity, cumulative energy, particle velocity, etc., or a combination of these with differing threshold levels? A full discussion of the effects of noise on fish is beyond the scope of this article, but can be found in recent reviews.[13,47]

[1]Department of the Environment and Water Resources (DEW), EPBC Act Policy Statement 2.1—Interaction between offshore seismic exploration and whales, Australian Government (2007).
[2]U.S. Environmental Protection Agency, "Small takes of marine mammals incidental to specified activities; seismic surveys in the Beaufort and Chukchi Seas off Alaska," Fed. Regist. **71**, 50027–50045 (2006).
[3]E. McCarthy, *International Regulation of Underwater Sounds: Establishing Rules and Standards to Address Ocean Noise Pollution* (Kluwer Academic, Boston, 2004).
[4]A. Engås, S. Løkkeborg, E. Ona, and A. V. Soldal, "Effects of seismic shooting on local abundance and catch rates of cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*)," Can. J. Fish. Aquat. Sci. **53**, 2238–2249 (1996).
[5]A. Engås and S. Løkkeborg, "Effects of seismic shooting and vessel-generated noise on fish behavior and catch rates," Bioacoustics **12**, 313–315 (2002).
[6]J. R. Skalski, W. H. Pearson, and C. I. Malme, "Effects of sounds from a geophysical survey device on catch-per-unit-effort in a hook-and-line fishery for rockfish (*Sebastes* ssp.)," Can. J. Fish. Aquat. Sci. **49**, 1357–1365 (1992).
[7]A. Slotte, K. Kansen, J. Dalen, and E. Ona, "Acoustic mapping of pelagic fish distribution and abundance in relation to a seismic shooting area off the Norwegian west coast," Fish. Res. **67**, 143–150 (2004).
[8]A. N. Popper, M. E. Smith, P. A. Cott, B. W. Hanna, A. O. MacGillivray, M. E. Austin, and D. A. Mann, "Effects of exposure to seismic airgun use on hearing of three fish species," J. Acoust. Soc. Am. **117**, 3958–3971 (2005).
[9]C. S. Wardle, T. J. Carter, G. G. Urquhart, A. D. F. Johnstone, A. M. Ziolkowski, G. Hampson, and D. Mackie, "Effects of seismic air guns on marine fish," Cont. Shelf Res. **21**, 1005–1027 (2001).
[10]J. P. Carricart-Ganivet, J. M. Lough, and D. J. Barnes, "Growth and luminescence characteristics in skeletons of massive Porites from a depth gradient in the central Great Barrier Reef," J. Exp. Mar. Biol. Ecol. **351**, 27–36 (2007).
[11]T. F. Cooper, S. Uthicke, C. Humpfhrey, and K. E. Fabricius, "Gradients in water column nutrients, sediment parameters, irradiance and coral reef development in the Whitsunday Region, central Great Barrier Reef," Estuarine Coastal Shelf Sci. **74**, 458–470 (2007).
[12]A. N. Popper, "Effects of anthropogenic sound on fishes," Fisheries **28**, 24–31 (2003).
[13]M. C. Hastings and A. N. Popper, "Effects of sound on fish," California Department of Transportation, Contract. No. 43A0139, 2005.
[14]J. T. Yelverton, D. R. Richmond, W. Hicks, K. Saunders, and E. R. Fletcher, "The relationship between fish size and their response to underwater blast," Report No. DNA 3677T, Defense Nuclear Agency, Washington, DC, 1975.
[15]J. H. Stuhmiller, K. H. H. Ho, M. J. Vander Vorst, K. T. Dodd, T. Fitzpatrick, and M. Mayorga, "A model of blast overpressure injury to the lung," J. Biomech. **29**, 227–234 (1996).
[16]R. D. McCauley, J. Fewtrell, and A. N. Popper, "High intensity anthropogenic sound damages fish ears," J. Acoust. Soc. Am. **113**, 638–642 (2003).
[17]J. J. Govoni, L. R. Settle, and M. A. West, "Trauma to juvenile pinfish and spot inflicted by submarine detonations," J. Aquat. Anim. Health **15**, 111–119 (2003).
[18]P. C. Etter, *Underwater Acoustic Modeling and Simulation*, 3rd ed. (Spon, London, 2003).
[19]R. D. McCauley, A. J. Duncan, J. Fewtrell, C. Jenner, M. Jenner, J. D. Penrose, R. I. T. Prince, A. Adhitya, J. Murdoch, and K. McCabe, "Marine seismic survey analysis and propagation of air-gun signals; and effects of air-gun exposure on humpback whales, sea turtles, fishes and squid," *Environmental Implications of Offshore Oil and Gas Development in Australia: Further Research* (Australian Petroleum Production and Exploration Association, Canberra, 2003), pp. 364–370, 381–385, and 498–521.
[20]ERT (Scotland) Ltd., "Third strategic environmental assessment for oil and gas activity in Ireland's offshore Atlantic waters: IOSEA3 Rockall Basin—Environmental report," submitted to the Department of Communications, Energy and Natural Resources Ireland (2003), available from http://www.dcenr.gov.ie/Natural/Petroleum+Affairs+Division/Irish+Offshore+Strategic+Environmental+Assessment+%28IOSEA+3%29/ (Last viewed January, 2009).
[21]Y. V. Gilchrest, T. N. Fetherston, and B. E. Neales, "Analysis of acoustic exposures on marine mammals for the proposed undersea warfare training range," NUWC Technical Report, Project No. C744088 (2008), available from http://projects.earthtech.com/uswtr/Library_index.htm (Last viewed

January, 2009).
[22]A. O. MacGillivray, "An acoustic modeling study of seismic airgun noise in Queen Charlotte basin," MS thesis, University of Victoria, BC, Canada (2006).
[23]A. Ziolkowski, "A Method for Calculating the Output Pressure Waveform from an Air Gun," Geophys. J. R. Astron. Soc. **21**, 137–161 (1970).
[24]W. H. Dragoset, "A comprehensive method for evaluating the design of airguns and airgun arrays," in 16th Annual Proceedings of the Offshore Technology Conference (1984), Vol. **3**, pp. 75–84.
[25]M. Laws, L. Hatton, and M. Haartsen, "Computer modeling of clustered airguns," First Break **8**, 331–338 (1990).
[26]M. Landro, "Modeling of GI gun signatures," Geophys. Prospect. **40**, 721–747 (1992).
[27]E. L. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[28]P. Kearey, M. Brooks, and I. Hill, *An Introduction to Geophysical Exploration* (Blackwell Scientific, Oxford, 2002).
[29]N. Barton, *Rock Quality, Seismic Velocity, Attenuation and Anisotropy* (Taylor & Francis, London, 2006).
[30]H. G. Brandes, A. J. Silva, and D. J. Walter, "Geo-acoustic characterization of calcareous seabed in the Florida Keys," Mar. Geol. **182**, 77–102 (2002).
[31]Y. Matsukura, A. Maekado, H. Aoki, T. Kogure, and Y. Kitano, "Surface lowering rates of uplifted limestone terraces estimated from the height of pedestals on a subtropical island of Japan," Earth Surf. Processes Landforms **32**, 1110–1115 (2007).
[32]S. S. Fu, C. Tao, M. Prasad, R. H. Wilkens, and L. N. Frazer, "Acoustic properties of coral sands, Waikiki, Hawaii," J. Acoust. Soc. Am. **115**, 2013–2020 (2004).
[33]J. Bowlin, J. Spiesberger, T. Duda, and L. Freitag, "Ocean acoustical ray-tracing software RAY," Technical Report No. WHOI-93-0, Woods Hole Oceanographic Institution, Woods Hole, MA, 1992.
[34]C. Erbe and D. M. Farmer, "A software model to estimate zones of impact on marine mammals around anthropogenic noise," J. Acoust. Soc. Am. **108**, 1327–1331 (2000).
[35]L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960).
[36]M. D. Collins, R. J. Cederberg, D. B. King, and S. A. Chin-Bing, "Comparison of algorithms for solving parabolic wave equations," J. Acoust. Soc. Am. **100**, 178–182 (1996).
[37]M. D. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[38]M. D. Collins, "The stabilized self-starter," J. Acoust. Soc. Am. **106**, 1724–1726 (1999).
[39]Z. Y. Zhang and C. T. Tindle, "Improved equivalent fluid approximations for a low shear speed ocean bottom," J. Acoust. Soc. Am. **98**, 3391–3396 (1995).
[40]J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, "SOM toolbox for Matlab 5," Report No. A57, SOM Toolbox Team, Helsinki University of Technology, PO Box 5400, FIN-02015 Hut, Finland, April 2000.
[41]J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, "Self-organizing map in Matlab: The SOM toolbox," in Proceedings of the Matlab DSP Conference, Espoo, Finland (1999), pp. 35–40.
[42]J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," IEEE Trans. Neural Netw. **11**, 586–600 (2000).
[43]X. Lurton, *An Introduction to Underwater Acoustics: Principles and Applications* (Springer, Berlin, 2002).
[44]M. D. Collins, "New and improved parabolic equation models," J. Acoust. Soc. Am. **104**, 1808 (1998).
[45]H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography* (Academic, Boston, 1998).
[46]M. E. Smith, A. S. Kane, and A. N. Popper, "Acoustical stress and hearing sensitivity in fishes: Does the linear threshold shift hypothesis hold water?," J. Exp. Biol. **207**, 3591–3602 (2004).
[47]A. N. Popper, "Effects of mid- and high-frequency sonars on fish," NUWC Technical Report, Contract N66604-07M-6056, available from http://projects.earthtech.com/uswtr/Library_index.htm (Last viewed January, 2009).

Erratum: "Born iterative reconstruction using perturbed-phase field estimates " [J. Acoust. Soc. Am., 124, 2353–2363 (2008)]

Jeffrey P. Astheimer
Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627

Robert C. Waag
Department of Electrical and Computer Engineering and Department of Imaging Sciences, University of Rochester, Rochester, New York 14627

(Received 30 January 2009; accepted 3 February 2009)
[DOI: 10.1121/1.3087431]

PACS number(s): 43.80.Ev, 43.60.Pt, 43.35.Wa, 43.80.Qf, 43.10.Vx

A printing error appears in Eqs. 6, 7, 8, 9, 10, 11, 14, 15 16, 17 20, 21, 40, 41, and 42. The two integral symbols in each of these equations were printed without a space between them and appear as $\iint$. The correct appearance of these double integrals is $\int\int$.

 0001-4966/2009/125(4)/2452/1/$25.00

ACOUSTICAL NEWS—USA

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

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

New Fellows of the Acoustical Society of America

Suzanne B. Boyce—For contributions to speech production and communication disorders

Reviewers of Manuscripts, 2008

Each year the Journal endeavors to publish a list of all the persons who reviewed manuscripts during the preceding year. Such a list is a compendium of names supplied by the Associate Editors. Because our peer review system depends strongly on the continuing anonymity of the reviewers, the Journal publishes these names in alphabetical order without identification of the Associate Editors who provided the names and without identification of the papers they reviewed. The primary reason for the publication of the list is to express the Journal's gratitude to its reviewers. Reviewing a paper is often a very time consuming and demanding task, and the anonymity requirement yields no professional recognition to those who generously provide their time to help the Associate Editors decide which papers should be published and to give constructive criticisms to the authors. The Journal is justifiably proud of this list, which includes a goodly proportion of all the researchers and eminent authorities in acoustics and related fields. In a compendium of this length, omissions and errors are inevitable. If anyone notices such, please send the corrections and missing names via e-mail or regular mail to either Elaine Moran (asa@aip.org) or Allan Pierce (adp@bu.edu).

Aasvang, Gunn Marit
Abascal, Ramon
Abbas, Paul J.
Abdala, Carolina
Abel, Sharon M.
Abom, Mats
Abraham, Douglas A.
Adachi, Seiji
Adams, Norman H.
Adank, Patti
Aguilar Soto, Natacha
Ahnert, Wolfgang
Ahroon, William A.
Ainslie, Michael A.
Ajdler, Thibaut
Akamatsu, Tomonari
Akay, Adnan
Akeroyd, Michael A.
Aktas, Murat K.
Albert, Donald G.
Alcantara, Jose I.
Alekseenko, Sergey
Aleshin, Vladislav
Alexander, Francis
Alexander, Jennifer A.
Alexander, Joshua M.
Alipour, Fariborz
Allard, Jean F.
Allen, John S.
Allen, Jont B.
Allen, Richard J.
Alpman, Emre
Altenberg, Evelyn
Amir, Noam
Amitay, Sygal
Ammi, Azzdine Y.
Anand, Akash
Anastasio, Mark A.
Anderson, Brian E.
Ando, Shigeru
Ando, Yoichi
Andre;-Obrecht, Regine
Andruski, Jean E.
Angster, Judit
Antila, Marko
Aoyama, Katsura
Apoux, Frederic C.
Aquino, Wilkins
Arai, Nobuaki
Aravamudhan, Radhika
Arnott, W. Pat
Aroyan, James
Arvaniti, Amalia
Arvelo, Juan
Arveson, Paul T.
Ashmead, Daniel H.
Askenfelt, Anders G.
Assmann, Peter F.
Astheimer, Jeffrey P.
Atalla, Noureddine
Atchley, Anthony A.
Atkins, Philip R.
Attenborough, Keith
Attias, Joseph
Au, Whitlow W.
Auer, Edward T.
Auregan, Yves

Bacon, Christophe
Bacon, Sid P.
Bader, Kenneth
Badin, Pierre
Baer, Thomas
Baer, Tom
Baik, Kyungmin
Bailey, Michael R.
Baillet, Laurent
Bailliet, Helene
Bailly, Christophe
Bailly, Gerard
Baker, Anne E.
Balakrishnan, Uma
Balasubramaniam, Krishnan
Ballandras, Sylvain J.
Ballato, Arthur
Bangs, Nathan
Banse, Rainer
Bao, Gang
Barbone, Paul E.
Barkmann, Reinhard
Barlow, Jay
Barney, Anna
Barrera-Figueroa, Salvador
Barron, Michael
Bashford, James A., Jr.
Baskent, Deniz
Bauer, Jay J.
Baulac, Marine
Bazua Duran, Carmen
Beauchamp, James
Beaujean, Pierre-Philippe J.
Bech, Soren
Becker, Kyle M.
Becot, FranHois-Xavier
Bedard, Alfred J.
Beddor, Patrice S.
Bednarik, Michal
Beedholm, Kristian
Behar, Alberto
Behne, Dawn M.
Beissner, Klaus
Benes, Ewald
Benetti, Massimiliano
Benki, Jose R.
Bennert, Thomas
Bent, Tessa C.
Benyahia, Lazhar
Beranek, Leo
Berchok, Catherine L.
Berengier, Michel
Berg, Bruce
Berg, Bruce G.
Berger, Elliott H.
Bergevin, Christopher
Berkhoff, Arthur
Bernard, Julien
Bernstein, Joshua G.
Bernstein, Leslie R.
Berry, Bernard
Berry, David A.
Berthelot, Yves H.
Berthommier, Frederic
Besing, Joan

 0001-4966/2009/125(4)/2453/8/$25.00

Best, Virginia
Betlehem, Terence
Beutelmann, Rainer
Bian, Lin
Biedron, Robert
Bierer, Julie A.
Bigelow, Timothy A.
Billon, Alexis
Bilsen, Frans A.
Birkett, Stephen
Bisgaard, Nikolai
Bishop, Garner C.
Bistafa, Sylvio R.
Bjor, Ole-Herman
Blacklock, Oliver S.
Blackwell, Susanna B.
Blamey, Peter J.
Blanc, Silvia
Bland, Roger W.
Blauert, Jens P.
Block, Frank
Bobrovnitskii, Yuri I.
Bocchieri, Enrico
Bockman, Alexander
Boe, Louis-Jean J.
Boemio, Anthony B
Boersma, Bendiks Jan
Boersma, Paul
Bomemann, Horst
Bond, Zinny S.
Boothroyd, Arthur
Bossy, Emmanuel
Bostrom, Anders
Botts, Jonathan
Bouakaz, Ayache
Boutin, Claude
Bowles, Ann E.
Boyce, Suzanne E.
Braasch, Jonas
Brachmanski, Stefan
Bradley, David T.
Bradley, John S.
Bradlow, Ann R.
Braida, Louis D.
Brambilla, Giovanni
Brand, Thomas
Brandes, T. S.
Branstetter, Brian K.
Bravo, Teresa
Breazeale, Mack A.
Breebaart, Jeroen
Bregman, Albert
Brennan, Michael J.
Brentner, Kenneth S.
Brewer, Peter G.
Bronkhorst, Adelbert W.
Brooks, Bennett M.
Brouwer, Harry H.
Brown, A. L.
Brown, Charles H.
Brown, David
Brown, Michael G.
Brown, Steven
Browning, David G.
Bruce, Ian C.
Brungart, Douglas
Brunskog, Jonas
Bryant, Gregory
Buck, John
Buckingham, Michael J.
Bucur, Voichita
Buder, Eugene H.
Bundgaard-Nielsen, Rikke L.
Bunnell, H. Timothy
Bunton, Kate E.
Burkard, Robert F.
Burnett, Theresa A.
Burnham, Denis K.
Burroughs, Courtney B.
Busby, Peter
Buss, Emily
Busse, Lawrence J.

Cabell, Randolph H.
Calamia, Paul T.
Calvisi, Michael
Calvo, David C.
Cameron, King
Campanella, Angelo J.
Campbell, D. M.
Campbell, Gregory S.
Candy, James V.
Cantrell, John H.
Capus, Chris G.
Carcaterra, Antonio
Carevic, Dragana
Carey, William M.
Carlile, Simon
Carlson, Thomas J.
Carlyon, Robert P.
Carney, Laurel
Caron, James
Carstensen, Edwin L.
Carter, G. Clifford
Castellini, Paolo
Castellote, Manuel
Castrejon-Pita, Alfonso A.
Castres, Fabrice O.
Catheline, Stefan
Cato, Douglas H.
Cavaco, Sofia
Cawley, Peter
Cazzolato, Ben S.
Chadwick, Richard
Chaigne, Antoine J.
Chalupper, Josef
Chambers, David H.
Chambers, James P.
Chambers, John
Chambers, Kyle
Chan, Roger W.
Chanaud, Robert C.
Chandrasekaran, Bharath
Chapman, David M.
Chapman, John
Chapman, N. Ross
Chappell, David J.
Chatellier, Ludovic
Chatterjee, Monita
Chen, Tianrun
Chen, Trevor H.
Chen, Yang
Cheng, Arthur C.
Chertoff, Mark
Chimenti, Dale E.
Chin-Bing, Stanley A.
Chiu, Ching-Sang
Cho, Taehong
Choi, Jee Woong
Cholewiak, Roger
Chordia, Parag
Chotiros, Nicholas P.
Chris, Plack
Chu, Dezhang
Chunchuzov, Igor
Chung, King
Church, Charles C.
Ciocca, Valter
Cipolla, Jeff
Clark, Cathy Ann
Clark, Christopher W.
Clemins, Patrick J.
Cleveland, Robin O.
Cleveland, Tom
Clopper, Cynthia G.
Cloutier, Guy
Cobbold, Richard S.
Cobo, Pedro
Colburn, H. Steven
Collet, Manuel
Collins, Leslie M.
Colonius, Tim
Conoir, Jean-Marc
Cooke, Martin
Cornuelle, Bruce D.
Corteel, Etienne
Cotoni, Vincent
Coussios, Constantin C.
Cowan, Robert S C.
Cox, Ben
Cox, Trevor J.
Cranford, Ted W.
Craster, Richard
Cristia, Alejandrina
Crocker, Daniel E.
Cross, Rod
Culling, John F.
Culver, Richard L.
Cummings, Alan
Cummins, Fred
Cunefare, Kenneth A.
Cusack, Rhodri
Cuschieri, Joseph M.
Cutanda Henriquez, Vicente

Dacol, Dalcio K.
Dahl, Peter H.
Dai, Huanping
Daigle, Gilles A.
Dalecki, Diane
D'Alessandro, Christophe R.
Dance, Stephen
Dancer, Armand L.
Dang, Jianwu
Darwin, Christopher J.
Dasgupta, Nilanjan
Dau, Torsten
David, Eddins
Davidson, Lisa
Davies, Patricia
Davila-Ross, Marina
Davis, Rickie R.
Dayton, Paul A.
Dazel, Olivier
Deane, Grant B.
De Boer, Bart
DeBruijn, Alex
De Bruijn, Werner
De Cheveigne, Alain
Declercq, Nico F.
De Coensel, Bert
Defrance, Guillaume
DeJong, Kenneth J.
De Jong, Nico
DelBalzo, Donald R.
Demany, Laurent
De Moustier, Christian P.
Dent, Micheal L.
Derode, Arnaud
De Rosa, Sergio
Desain, Peter
Desharnais, Francine
Deutsch, Diana
Devillers, Laurence Y.
De Vries, Diemer
Dhar, Sumitrajit
Di, Xiao E.
DiCanio, Christian T.
Dickey, Joe W.
Dilley, Laura C.
Dillier, Norbert
Dillon, Harvey A.
Ding, Desheng
Ding, Heping
Diperna, Daniel
Disterhoft, John
Divenyi, Pierre L.
Dizon, Roberto M.
Djendi, Mohamed
Dobie, Robert
Dodd, Tony
Doherty, Karen A.
Doinikov, Alexander A.
Don, Manuel
Donaldson, Gail S.
Donskoy, Dimitri M.
Dooling, Robert J.
Dorman, Michael F.
Dosso, Stan E.
Dougherty, Robert
Doval, Boris
Dowling, David R.
Dowling, Jay
Dreisbach, Laura E.
Drennan, Ward R.
Dreschler, Wouter A.

Drinkwater, Bruce W.
Drullman, Rob
Druzhinin, Grigory
D'Spain, Gerald L.
Dubno, Judy R.
Dubus, Bertrand B.
Duda, Richard O.
Duda, Timothy F.
Duennebier, Fred
Duifhuis, Hendrikus
Durlach, Nathaniel I.
Dushaw, Brian D.
Dvorkin, Jack
Dye, Raymond H.
Dzieciuch, Matthew A.
Dzikowicz, Benjamin R.

Eadie, Tanya L.
Ebenezer, D. D.
Economou, Eleftherios
Eddins, David A.
Edds-Walton, Peggy
Edelmann, Geoffrey F.
Edmonds, Barrie A.
Eisenberg, Laurie S.
Elberling, Claus
Ellermeier, Wolfgang
Elliott, Stephen J.
Ellis, Christien
Ellis, Daniel P.
Emelianov, Stanislav Y.
Engwall, Olov
Epstein, Michael J.
Erath, Byron D.
Erbe, Christine
Erickson, Molly
Erofeev, Vladimir
Escudero, Paola
Espinosa, Victor
Evans, Richard B.
Everbach, Erich C.
Eversman, Walter
Everstein, Gordon
Ewert, Stephan D.

Faber, Alice
Fabre, Benoit
Fahey, Paul F.
Fahnline, John B.
Fahy, Frank
Falcone, Mauro
Farassat, F.
Farina, Angelo
Fastl, Hugo
Faulkner, Andrew
Fawcett, John A.
Fay, Richard R.
Fazi, Filippo M.
Fechter, Larry
Feijoo, Gonzalo R.
Feit, David
Fellah, Zine el Abiddine
Fenton, Brock
Ferguson, Sarah H.
Festen, Joost M.
Fialkowski, Laurie T.
Fields, James M.
Filipanits, Frank
Fine, Michael L.
Finette, Steven I.
Finlayson, Paul
Finneran, James J.
Fisher, Karl A.
Fitch, W. Tecumseh T.
Fitzgerald, Tracy S.
Fitzpatrick, John A.
Flege, James E.
Fletcher, Neville H.
Fligor, Brian
Florentine, Mary
Folkins, John W.
Folkow, Peter D.
Fon, Janice
Foote, Kenneth G.
Formby, Craig
Forssen, Jens
Fotiadis, Dimitrios I.
Fourakis, Margaritis S.
Fowlkes, Jeffrey B.
Fox, Robert A.
Francis, David T.
Frank, Austin F.
Frank, Scott D.
Frankel, Steven H.
Frankenthal, Shimshon
Franzoni, Linda P.
Fredouille, Corinne
Freed, Daniel J.
Freitag, Lee
Friesen, Lendra
Friis, Lars
Friot, Emmanuel
Frisina, Robert D.
Fritz, Claudia
Fromm, David M.
Fromme, Paul
Fu, Qian-Jie
Fujimura, Osamu
Fulcher, Lewis
Funnell, W. Robert J.
Furihata, Kenji
Furst, Miriam
Furukawa, Shigeto

Gade, Anders Christian S.
Gaeta, Rick
Galeotti, Paolo
Galland, M.A.
Gallun, Frederick J.
Gan, Rong
Gannot, Sharon
Garai, Massimo
Garcia-Lecumberri, Maria-Luisa L.
Gardonio, Paolo
Garnier, Vincent
Garrelick, Joel M.
Gaumond, Charles F.
Gaunaurd, Guillermo C.
Gauthier, Claude
Gauthier, Philippe-Aubert
Gautier, Francois
Gee, Kent L.
Geller, Drew A.
Gendron, Paul
Geng, Christian
Genuit, Klaus
Gerhold, Carl H.
Gerratt, Bruce R.
Gerretsen, Eddy
Gerstoft, Peter
Geurts, Luc
Ghazanfar, Asif
Ghose, Kaushik
Ghosh Roy, Dilip N.
Ghoshal, Goutam
Gibiat, Vincent
Giguere, Christian
Gilbert, Joel
Gillbert, Ken
Gillespie, Douglas M.
Ginsberg, Jerry H.
Giordano, Bruno L.
Giurgiutiu, Victor
Givoli, Dan
Gjerdingen, Robert
Glasberg, Brian R.
Gleich, Otto
Glorieux, Christ
Gockel, Hedwig E.
Goddard, Gregory
Godin, Oleg A.
Goebl, Werner
Goedecke, George H.
Goffman, Lisa
Goldinger, Stephen
Goldsworthy, Ray L.
Gomes, Jesper
Gomes, Joao P.
Gordon, Grant
Gordon, Jonathan C.
Gordon-Salant, Sandra
Gorga, Michael P.
Gottfried, Terry L.
Gough, Colin
Goupell, Matthew J.
Gow, David W.
Gramtorp, Johan
Grantham, D. W.
Graziani, Giorgio
Green, Dale
Green, Phil
Green, Tim
Greenberg, Steven
Greene, Charles R., Jr.
Greenleaf, James F.
Gregory, Michelle L.
Grieco-Calub, Tina
Griefahn, Barbara
Griesinger, David H.
Griffa, Michele
Griffiths, Timothy D.
Grimault, Nicolas
Grinchenko, Victor T.
Grose, John H.
Grosh, Karl
Guillot, Francois M.
Guinan, John J., Jr.
Guion, Susan G.
Gumerov, Nail A.
Gupta, Prahlad
Gusev, Vitaly
Guski, Rainer
Gutschalk, Alexander
Guven, Emre
Guzina, Bojan
Gygi, Brian M.

Hübelt, Jorn
Haberman, Michael R.
Haberman, Robert C.
Hachiya, Hiroyuki
Hahn, Mariah S.
Hahn, Thomas R.
Haiat, Guillaume
Hain, Timothy C.
Hall, Deborah
Hall, Joseph W.
Hall, Neal A.
Hall, Joseph W., III
HallJ, Pierre A.
Halpern, Andrea
Hamdi, Mohamed-Ali
Hamernik, Roger P.
Hamilton, Mark F.
Hammershoi, Dorte
Handegard, Nils Olav
Hanes, Peter
Hanford, Amanda D.
Hansen, Colin H.
Hansen, John H.
Hansen, Uwe J.
Hanson, Helen M.
Hanyga, Andrzej
Hardin, Jay C.
Harma, Aki S.
Harnsberger, James D.
Harris, Gerald R.
Harrison, Christopher H.
Harrison, Robert V.
Hartmann, William M.
Hartung, Klaus
Hasegawa-Johnson, Mark A.
Hastings, Mardi C.
Hawkins, Anthony D.
Hawley, Monica L.
Hay, Alex E.
Hayek, Sabih I.
Hayes-Harb, Rachel L.
Hazan, Valerie
He, Zhengyao
Heaney, Kevin D.
Hedberg, Claes M.
Heffner, Henry E.
Hefner, Brian T.
Heide-Jorgensen, Mads Peter
Heil, Peter

Heinz, Michael G.
Helfer, Karen S.
Heller, Laurie M.
Hellweg, Robert
Henderson, Brenda S.
Henyey, Frank S.
Herd, Wendy
Herdic, Peter C.
Hermand, Jean-Pierre
Hermansky, Hynek
Hermes, Dik J.
Hershey, John R.
Herzel, Hanspeter
Herzing, Denise
Herzog, Philippe
Hesthaven, Jan
Hickey, Craig J.
Hicks, Candace
Hidaka, Takayuki
Higgs, Dennis M.
Hilgenfeldt, Sascha
Hillenbrand, James M.
Hiller, Robert A.
Hinders, Mark K.
Hinich, Melvin J.
Hirata, Yukari
Hirose, Keikichi
Hirschberg, Avraham
Hirsekorn, Sigrun
Hirshberg, Avraham
Hiryu, Shizuko
Hodge, Megan M.
Hodgkiss, William S.
Hodgson, Murray R.
Hoffmann-Kuhnt, Matthias
Hohmann, Volker
Holland, Charles
Holland, Nicholas V., III
Hollien, Harry
Holm, Sverre
Holmes, Jason D.
Holt, Marla M.
Holt, R. G.
Holube, Inga
Holzfuss, Joachim
Honda, Kiyoshi
Hoole, Philip A.
Hoover, K Anthony
Hopgood, James R.
Hopkins, Carl
Hopkins, Kathryn
Horacek, Jaromir
Horne, John K.
Hornsby, Benjamin W.
Horoshenkov, Kirill V.
Horwitz, Amy R.
Hosokawa, Atsushi
Hosten, Bernard
Houde, John F.
Houle, Phil
House, Jill
Houser, Dorian S.
Houtgast, Tammo
Houtsma, Adrian
Howard, David M.
Howe, Bruce M.
Howe, Michael S.
Hsiao, Steven
Hu, Xiangyu
Huang, Jui-Ting
Hume, Ken I.
Humphrey, Victor F.
Hunter, Eric J.
Huntley Bahr, Ruth A.
Hursky, Paul
Hussain, Zahir M.
Hustad, Katherine
Huybregts, Neil

Idsardi, Bill
Ihlefeld, Antje
Ihlenburg, Frank
Ilinskii, Yurii A.
Insana, Michael F.
Ioup, George E.
Iskarous, Khalil
Ison, James R.

Jackson, Darrell R.
Jackson, Philip J. B.
Jacobs, Laurence J.
Jacobsen, Finn
Jaeger, Stephen M.
Janata, Petr
Jang, Seung-Ho
Janssen, Thomas
Jansson, Erik V.
Jansson, Tomas
Jaouen, Luc
Jarzynski, Jacek
Jean, Philippe
Jelinek, Milan
Jeon, Jin Yong
Jessen, Michael
Jesteadt, Walt
Jesus, Sergio M.
Jiang, Yongmin
Jin, Craig
Jing, Xiaodong
Jing, Yun
Job, Soames
Johansson, Torbjorn A.
John, Rosowski
Johnson, David L.
Johnson, Don H.
Johnson, Mark P.
Johnstone, Patti M.
Jones, Gary L.
Jones, Jeffery A.
Jongman, Allard
Jumars, Peter

Kabak, B.
Kaduchak, Gregory
Kain, Alexander
Kaliski, Kenneth H.
Kalluri, Radha
Kamm, Teresa M.
Kang, Jian
Kapralos, Bill
Karjalainen, Matti A.
Karlsson, Inger
Katz, Brian F.
Kaufman, Jonathan
Kausel, Wilfried
Kawahara, Hideki
Kazanina, Nina
Keefe, Douglas H.
Keith, William
Kellermann, Walter L.
Kelly, James F.
Kelly, Jonathan W.
Kendall, Roger
Keolian, Robert M.
Kergomard, Jean
Kewley-Port, Diane
Khismatullin, Damir B.
Khokhlova, Vera A.
Khosla, Sid
Kidd, Gary R.
Kidd, Gerald, Jr.
Kiefte, Michael
Kilfoyle, Daniel
Kim, Hyunsoon
Kim, Jaehwan
Kim, Yang-Hann
Kim, Yoon Young
Kimura, Masao
Kirby, Raymond
Kirsteins, Ivars
Kistler, Doris J.
Klaboe, Ronny
Klapuri, Anssi
Klemuk, Sarah A.
Klippel, Wolfgang
Klos, Jacob
Kluender, Keith R.
Klump, Georg M.
Kob, Malte
Kobler, James B.
Koch, Robert A.
Kochanski, Greg P.
Koehnke, Janet D.
Koenig, Laura L.
Kohler, Klaus J.
Kohlrausch, Armin
Koike, Takuji
Kolaini, Ali R.
Kolios, Michael C.
Kollmeier, Birger
Kong, Eunjong
Kong, Ying-Yee
Konrad-Martin, Dawn
Kopco, Norbert
Korman, Murray S.
Korneliussen, Rolf J.
Krahe, Detlef
Kral, Peter
Krause, Jean
Kreider, Wayne
Kreiman, Jody E.
Kripfgans, Oliver D.
Kristensson, Gerhard
Kristiansen, Ulf R.
Krothapalli, Anjaneyulu
Krumbholz, Katrin
Krylov, Victor V.
Kucinschi, Bogdan R.
Kundu, Tribikram
Kuperman, William A.
Kuster, Martin
Kuttruff, K. Heinrich
Kuznetsova, Larisa
Kwon, Oh-Wook
Kyhn, Line A.

Laback, Bernhard
Lacefield, James C.
Lachs, Lorin
Ladd, Robert
Ladich, Friedrich
Lafarge, Denis
Lafon, Cyril
Lagree, Pierre-Yves
Lam, Yiu W.
Lammers, Marc O.
Lamont, Alexandra
Lampropoulos, Iraklis E.
Landsberger, Brian J.
Landsberger, David M.
Langley, Robin S.
Langner, Gerald
Langton, Christian
Lanza di Scalea, Francesco
Lapsley Miller, Judi
Larose, Eric
Larson, Charles R.
Larson, Gregg D.
Lau, Siu-Kit
Lauer, Amanda M.
Laugier, Pascal
Laukka, Petri
Lauriks, Walter
Lauterborn, Werner Horst
Lavandier, Catherine
Lavandier, Mathieu
Lavery, Andone C.
Lawrenson, Chris
Lawson, Gareth
Leach, William M.
Le Bas, Pierre-Yves
Lebourges-Dhaussy, Anne
Leclaire, Philippe
Lee, Jungmee
Lee, Moohyung
Lee, Soo-Young
Lee, Tan
Leek, Marjorie R.
Lehman, Sean K.
Leibold, Lori J.
Leighton, Tim
Leighton, Timothy G.
Lentz, Jennifer J.
LePage, Kevin D.
Lerch, Reinhard
Lercher, Peter
Lesniewski, Peter J.

Lethiecq, Marc
Letowski, Tomasz R.
LeVeque, Randall
Levi, Susannah V.
Levitin, Daniel
Levitt, Harry
Levy, Erika
Lewy, Serge
Li, Kai Ming M.
Li, Songhai
Li, Weichang
Li, XinXin
Li Hegner, Yiwen
Lieberman, D.
Lin, Tian Ran
Lin, Wei
Lineton, Ben
Lingevitch, Joseph F.
Lipkens, Bart
Lissenden, Cliff
Litovsky, Ruth Y.
Littrell, Robert J.
Litvak, Leonid M.
Liu, Chang
Liu, Fei
Liu, Gui Rong
Liu, Hanjun
Liu, Qing-Huo
Liu, Sheng
Liu, Zunping
Lobdell, Bryce E.
Lobel, Philip
Lobkis, Oleg I.
LoCasto, Paul C.
Loebach, Jeremy L.
Lofqvist, Anders
Loizou, Philipos C.
Lokki, Tapio
Long, Christopher J.
Long, Glenis R.
Long, Lyle N.
Lonsbury-Martin, Brenda L.
Lopez-Poveda, Enrique A.
Lorenzi, Christian
Lotto, Andrew J.
Lotton, Pierrick
Loughlin, Patrick
Louisnard, Olivier
Lovell, Jonathan
Lowe, Michael J.
Lu, Ching-Ta
Lu, John
Lu, Tom
Lubbers, Jaap
Lubman, David
Lucero, Jorge C.
Lugli, Marco
Lukashkin, Andrei N.
Luo, Xin
Lutfi, Robert A.
Lutkenhoner, Bernd
Lutman, Mark E.
Lyon, Richard H.
Lyzenga, Johannes

Macaskill, Charlie
Mace, Brian R.
Macherey, Olivier
Macmillan, Neil
Macpherson, Ewan A.
Madsen, Peter T.
Maekawa, Kikuo
Magnuson, James S.
Mainardi, Francesco
Majdak, Piotr
Makov, Yuri
Makris, Nicholas C.
Maldague, Xavier
Mamou, Jonathan
Mann, David A.
Marburg, Steffen
Marengo, Edwin A.
Margerin, Ludovic
Marshall, Bill
Marshall, Lynne
Marston, Philip L.
Martens, Jean-Pierre
Martin, Glen
Martin, Philippe
Martinez, Rudolph
Maruvada, Subha
Maschke, Christian
Mason, Christine R.
Masson, Patrice
Mast, T. Douglas
Masters, William M.
Mathevon, Nicolas
Matsuo, Ikuo
Mattys, Sven L.
Mauermann, Manfred
Maury, Dominique
Max, Ludo
Maynard, Julian D.
McAdams, Stephen E.
McAlpine, David
McCauley, Rob
McCoy, John J.
McCracken, Gary
McDaniel, James G.
McDannold, Nathan
McDonald, Bryant E.
McDonald, Mark A.
McFadden, Dennis
McFadden, Sandra L.
McGough, Robert J.
McGowan, Richard S.
McGuire, Grant L.
McKay, Colette M.
McKenna, Mihan H.
McLaughlin, Stephen
McLean, James S.
McPherson, Bradley
McRoberts, Gerald
Means, Steven L.
Mecklenbrauker, Christoph F.
Meddis, Ray
Mehl, James B.
Mehler, Jacques
Meinke, Deanna K.

Mellert, Volker
Mellinger, David K.
Mercado, Eduardo, III
Meunier, Sabine
Meyer, Jens M.
Meyer-Baese, Anke
Michaels, Thomas E.
Michalopoulou, Zoi-Heleni
Michaud, David
Micheyl, Christophe
Middlebrooks, John C.
Miele, Joshua
Migliori, Albert
Mignerey, Peter C.
Miksis-Olds, Jennifer L.
Miller, Douglas L.
Miller, James
Miller, James G.
Miller, James H.
Miller, Lee A.
Milner, Ben P.
Milton, Graeme
Mironov, Mikhail A.
Mishchenko, Michael
Miskiewicz, Andrzej J.
Misra, Anil
Mitri, Farid
Miyoshi, Masato
Möbius, Bernd
Mobley, Joel
Moffett, Mark B.
Möhler, Ulrich
Moilanen, Petro
Mokhtari, Parham
Moldover, Michael R.
Moleti, Arturo
Monahan, Edward C.
Mongeau, Luc G.
Monson, Brian B.
Mooney, T. Aran
Moore, Brian C.
Moore, David R.
Moore, Elliot
Moorhouse, Andy T.
Moreau, Andre
Moretti, David J.
Morozov, Andrey K.
Morris, Philip J.
Morrison, Andrew
Morse, Scot F.
Mortessagne, Fabrice
Moss, Cynthia F.
Mossbridge, Julia A.
Mosto, Patricia
Mould, John
Mountain, David C.
Mourjopoulos, John
Muehleisen, Ralph T.
Muesch, Hannes
Muir, Thomas G.
Muller, Rolf
Munhall, Kevin G.
Munjal, M. L.
Munson, Benjamin R.

Murphy, Dana
Murphy, William J.
Murray, Todd W.
Myers, Matthew R.

Naghshineh, Koorosh
Nagy, Peter B.
Nakatani, Tomohiro
Nark, Douglas
Naylor, Graham
Nearey, Terrance M.
Nederveen, Cornelis J.
Neely, Stephen T.
Neff, Donna L.
Neild, Adrian
Nelken, Israel
Nelson, Peggy B.
Nemeth, Erwin
Neubauer, Juergen
Newhall, Bruce K.
Ng, Manwa L.
Niebuhr, Oliver
Nielsen, Peter L.
Nielsen, Traci
Niezrecki, Christopher
Nightingale, K. R.
Nilsson, Mats E.
Nip, Ignatius
Nishi, Kanae
Nissen, Shawn L.
Nittrouer, Susan N.
Nobile, Matthew A.
Noble, William
Nocke, Christian
Nolte, Loren W.
Nordholm, Sven
Norris, Andrew N.
Norris, David E.
Norris, Dennis
Norris, Thomas F.
Norton, Guy V.
Nosal, Eva-Marie
Nowak, Robert
Nuttall, Alfred L.
Nyborg, Wesley L.
Nygaard, Lynne C.

O'Brien, William D., Jr.
Obrist, Martin K.
Ochmann, Martin
Odom, Robert I.
O'Donnell, Matthew
Oelze, Michael L.
Oerlemans, Stefan
Oglesbee, Eric N.
Ohayon, Roger
Ohde, Ralph N.
Ohga, Juro
Ohl, Claus-Dieter
Oldham, David J.
Oleson, Erin M.
Ollivier, Sébastien
Olsen, John
Olson, Elizabeth S.
Omoto, Akira

Ona, Egil
Oppenheim, Irving J.
O'Regan, Stephen D.
Orris, Gregory J.
Ostrovsky, Lev A.
Oswald, Julie N.
Otsuru, Toru
Outing, Donald A.
Ouyang, Huajiang
Owren, Michael J.
Oxenham, Andrew J.
Ozdamar, Ozcan

Padilla, Frederic
Page, John H.
Pagneux, Vincent
Palmer, Alan R.
Paltauf, Guenther
Paolero, Anthony E.
Papanicolaou, George C.
Parrott, Tony L.
Parthasarathi, Anand
Pascal, Jean-Claude
Pasqualini, Donatella
Pastore, Richard E.
Patil, Abhay
Patterson, Douglas
Patterson, Roy D.
Peat, Keith S.
Pecorari, Claudio V.
Pedersen, Peder C.
Pelekasis, Nikos A.
Pelorson, Xavier
Peng, Shu-Chen
Perrin, Robert
Perrot, Camille
Persson Waye, Kerstin
Petculsecu, Gabriela
Peter, Gerstoft
Peters, Francois
Peters, Robert W.
Peterson, Michael L.
Petersson, Bjrrn A.
Petrov, Viacheslav
Pfingst, Bryan E.
Pfordresher, Peter Q.
Phatak, Sandeep A.
Phillips, Dennis P.
Photiadis, Douglas
Picaut, JudicaJl
Picton, Terence W.
Pierce, Allan D.
Pietrzko, Stanislaw
Pinton, Gianmarco F.
Piquette, Jean C.
Pishchalnikov, Yuri A.
Pispola, Giulio
Pitermann, Michel
Pitt, Mark A.
Piwakowski, Bogdan
Plack, Christopher J.
Plesniak, Michael
Poeppel, David
Poerschmann, Christoph
Poese, Matthew E.
Poissant, Sarah F.
Polack, Jean-Dominique
Polka, Linda
Pols, Louis C.
Pompei, F. Joseph
Poole, Joyce H.
Popelka, Gerald R.
Popper, Arthur N.
Porter, Michael B.
Posey, Joe W.
Post, John T.
Potapov, Aleksandr
Poterek, Thomas J.
Potter, John
Potty, Gopu R.
Poulsen, Torben
Prada, Claire
Prager, Jens
Prasanna, S.R.M.
Pratap, Rudra
Preis, Anna
Preisig, James C.
Premus, Vincent
Pressnitzer, Daniel
Price, G. R.
Prieve, Beth A.
Prodi, Nicola
Prosperetti, Andrea
Protopapas, Athanassios
Protopappas, Vasilios C.
Pulkki, Ville
Purcell, David W.
Puria, Sunil

Qian-Jie, Fu
Qin, Michael K.
Qin, Yi-Xian
Qu, Jianmin
Quatieri, Thomas F.
Quené, Hugo
Quijano, Jorge E.
Quinlan, Angela E.

Rabbitt, Richard
Radcliffe, Clark
Radziemski, Leon
Rafaely, Boaz
Rajan, Subrahaniam D.
Rakerd, Brad S.
Rakowski, Andrzej
Ramsay, Gordon J.
Ramus, Franck
Rank, Erhard
Rankin, Shannon
Rao, Preeti
Rasmussen, Arne N.
Raspet, Richard
Rathsam, Jonathan
Ravicz, Michael E.
Rayess, Nassif E.
Recasens, Daniel
Reed, Charlotte M.
Reichmuth, Colleen
Reid, John M.
Reidenberg, Joy
Remez, Robert E.
Ren, Tianying
Repp, Bruno H.
Reuter, Karen
Rhebergen, Koenraad S.
Richard, McGowan S.
Richards, Virginia M.
Richardson, Michael D.
Richtsmeier, Peter
Riede, Tobias
Rienstra, Sjoerd W.
Rindel, Jens H.
Riquimaroux, Hiroshi
Rivens, Ian
Roberts, Dale
Roch, Marie
Rockwell, Patricia
Roengpitya, Rungpat
Roger, Michel
Rogers, Catherine L.
Rogers, Peter H.
Roh, Heui-Seol
Rokhlin, Stanislav I.
Roman, Nicoleta
Roosnek, Nico
Rosowski, John J.
Rossing, Thomas D.
Roumeliotis, John A.
Rountree, Rodney
Rouseff, Daniel
Roux, Philippe
Roy, Kenneth P.
Royer, Daniel
Royston, Thomas J.
Rozanov, Konstantin
Rubinstein, Jay
Ruggero, Mario A.
Russell, Ian J.
Russell, Michael K.
Ruzzene, Massimo
Ryan, James G.
Rypina, Irina I.
Ryzhik, Lenya

Saberi, Kourosh
Sabra, Karim G.
Sadler, Brian M.
Sahin, Zekeriyya
Saijyou, Kenji
Samarra, Filipa I.
Samuel, Arthur G.
Sanchez-Morcillo, Victor J.
Sanders, Lisa D.
Sangani, Ashok
Sankin, Georgy
Santillan, Arturo O.
Saoji, Aniket A.
Sapozhnikov, Oleg A.
Sarkar, Kausik
Sarkissian, Angie
Sarradj, Ennes
Sarvas, Jukka
Sarvazyan, Armen
Sato, Hiroshi
Satyanarayan, Louis
Savin, Eric M.
Sawusch, James R.
Sboros, Vassilis
Scalerandi, Marco
Scandrett, Clyde L.
Scarborough, Rebecca
Schafer, Mark E.
Schairer, Kim S.
Scharenborg, Odette
Scharf, Bertram
Scharinger, Matthias
Scheifele, Peter M.
Scherer, Ronald C.
Schlauch, Robert S.
Schmidlin, Dean J.
Schmiedt, Richard A.
Schmitt, Denis P.
Schmitt, Douglas R.
Schmitz, Chris
Schneider, Bruce A.
Schneider, John B.
Schneider, Stefan
Schnupp, Jan
Schoentgen, Jean B.
Scholik-Schlomer, Amy
Scholte, Rick
Schomer, Paul D.
Schreckenberg, Dirk
Schreiner, Christoph E.
Schuemer, Rudolf
Schwartz, Jean-Luc
Schwartz, Joshua
Schwarz, Raphael
Schwenke, Roger W.
Scott, Sophie K.
Scott, Waymond R.
Seeber, Bernhard U.
Seidenberg, Mark
Seiner, Jack
Sek, Aleksander P.
Selamet, Ahmet
Selouani, Sid-Ahmed
Semblat, Jean-Francois
Semmlow, John L.
Serafin, Stefania
Serniclaes, Willy
Serpanos, Yula
Sessler, Gerhard M.
Sethares, William A.
Settles, Gary S.
Seybert, Andrew F.
Sgard, Franck C.
Shackleton, Trevor M.
Shafiro, Valeriy
Shamma, Shihab A.
Shannon, Robert V.
Shaw, Adam
Sheft, Stanley
Shepard, W S.
Shera, Christopher A.
Shiller, Douglas M.
Shimizu, Yasushi

Shinn-Cunningham, Barbara G.
Shirron, Joseph J.
Shore, Susan E.
Shorter, Philip J.
Shrivastav, Rahul
Shub, Daniel E.
Shung, K K.
Sicuranza, Giovanni L.
Siderius, Martin
Siegel, Jonathan H.
Sijtsma, Pieter
Siltanen, Samuel A.
Silva, Antonio
Silva, David J.
Silva, Glauber T.
Simmons, Andrea M.
Simon, Malene
Simonetti, Francesco
Simpson, Brian D.
Simpson, Roger
Sinex, Donal G.
Singh, Leher
Singh, Satish
Sinkus, Ralph
Sinnott, Joan M.
Sisneros, Joseph A.
Skarnitzl, Radek
Skarsoulis, Emmanuel K.
Skowronski, Mark D.
Slaton, William V.
Slifka, Janet L.
Slowiaczek, Louisa M.
Smeulders, David
Smith, Bruce L.
Smith, Caroline L.
Smith, John
Smith, John R.
Smith, Julius O.
Smith, Kevin B.
Smith, Nadine B.
Smith, Zachary M.
Smotherman, Michael S.
Snellen, Mirjam
Snieder, Roel
Snowdon, Charles
Snyder, Joel S.
Soeta, Yoshiharu
Soize, Christian
Solna, Knut
Solodov, Igor
Solomon, Nancy P.
Soltis, Joseph
Sommers, Mitchell S.
Song, Hee-Chun
Sorokin, Sergey
Sousa-Lima, Renata S.
Souza, Luiz L.
Souza, Pamela E.
Spehar, Brent
Spicer, James B.
Spiesberger, John L.
Spoor, Philip S.
Sprofera, Joseph D.
Srinivasan, Soundararajan
Stafford, Kathleen M.
Stainsby, Thomas H.
Stallen, Pieter J.
Stanzial, Domenico
Stathopoulos, Elaine T.
Stecker, G. Christopher
Steinschneider, Mitchell
Stellmack, Mark A.
Stelmachowicz, Patricia G.
Stemberger, Joseph P.
Stenfelt, Stefan P.
Stimpert, Alison
Stinson, Michael R.
Stone, Maureen L.
Stone, Michael A.
Story, Brad H.
Strain, George M.
Strange, Winifred
Strickland, Elizabeth A.
Stride, Eleanor
Studebaker, Gerald A.
Stulov, Anatoli
Sturm, F.
Stylianou, Yannis
Sugimoto, Nobumasa
Sukittanon, Somsak
Summerfield, Arthur Q.
Summerfield, Quentin
Summers, Jason E.
Summers, Van
Sumner, Chris
Sun, Xiaofeng
Sundaram, Shiva
Sundberg, Johan E.
Supin, Alexander Y.
Surlykke, Annemarie
Sutherland, Louis C.
Sutin, Alexander
Svec, Jan G.
Svirsky, Mario
Swearingen, Michelle E.
Swift, Gregory W.
Syrdal, Ann K.
Szabo, Thomas L.
Szeri, Andrew J.
Szuberla, Curt A.

Tabain, Marija
Tachibana, Hideki
Takeda, M.W.
Takemoto, Hironori
Tallal, Paula
Talmadge, Carrick L.
Talmant, Marylin
Tan, Bernard T.
Tan, Chee Wee
Tang, Dajun
Tanter, Mickael
Tarnow, Viggo
Taroudakis, Michael
Tasko, Stephen M.
Tatarinov, Alexey
Ten Bosch, Louis
TenCate, James A.
Ter Haar, Gail
Ternstrom, Sten
Tesei, Alessandra
Thibeault, Susan
Thode, Aaron M.
Thomas, Charles
Thomas, Jean-Hugh
Thomas, Jean-Louis
Thompson, David J
Thompson, Eric R.
Thompson, Stephen C.
Thomson, Scott L.
Thorne, Peter D.
Thornely-Taylor, Rupert M.
Thorsos, Eric I.
Throckmorton, Chandra S.
Tian, Yi
Tiede, Mark K.
Tindle, Christopher T.
Tiseo, Barbara
Titze, Ingo R.
Todd, Neil P.
Tokarev, Vadim
Tokuda, Isao T.
Tollefsen, Dag
Tollin, Daniel J.
Torre, Peter
Tougaard, Jakob
Tourin, Arnaud
Traunmuller, Hartmut
Traykovski, Peter
Treeby, Bradley E.
Trevorrow, Mark V.
Treyssede, Fabien
Trudeau, Micahel
Tsao, Ying-Chiao
Tsogka, Chrysoula
Tubis, Arnold
Tucker, Benjamin V.
Tufts, Jennifer B.
Tun, Pat
Tuomainen, Jyrki
Turgut, Altan
Turner, Christopher W.
Turner, Joseph A.
Tyack, Peter L.
Tyagi, Manav
Tye-Murray, Nancy
Tzanetakis, George

Umesh, S.
Umnova, Olga
Urban, Matthew W.

Vagle, Svein
Valiere, Jean-Christophe
Valimaki, Vesa
Van Alphen, Petra M.
Van de Par, Steven
Van Den Abeele, Koen
Van Dijk, Pim
Vander Meulen, Francois
Vander Werff, Kathy
Van Heuven, Vincent J.
Van Hoesel, Richard J.
Van Netten, Sietse M.
Van Parijs, Sofie M.
Van Renterghem, Timothy
Van Walree, Paul A.
Van Wieringen, Astrid
Varadan, Vasundara
Varslot, Trond
Vaseghi, Saeed
Vavrycuk, Vaclav
Vecherin, Sergey
Vecherin, Sergey N.
Veijola, Timo
Vendittis, David J.
Vercammen, Martijn
Verfuß, Ursula K.
Verhey, Jesko L.
Vermeir, Gerrit L.
Versluis, Michel
Vieira, Jose
Vignola, Joseph F.
Vignon, Francois
Virieux, Jean
Virovlyansky, Anatoly L.
Vogt, Michael
Von Wagner, Utz
Vongpaisal, Tara
Vorlaender, Michael
Voronovich, Alexander G.
Vorperian, Houri K.
Vos, Joos
Voss, Susan E.

Wade, Travis
Wagner, Anita E.
Wagner, Randall P.
Wagstaff, Ronald A.
Wahlberg, Magnus
Wakeland, Ray S.
Wallace, Kirk D.
Wang, Chao
Wang, DeLiang
Wang, Hsin-Min
Wang, Lijuan
Wang, Lily M.
Wang, Lugen
Wang, Meng
Wang, Yue
Wang, Yue-Sheng
Ward, David
Warren, Joseph D.
Washburn, Karl B.
Waterhouse, Richard V.
Waters, Kendall R.
Waters, Zachary J.
Watkins, Anthony J.
Waxler, Roger M.
Wayland, Ratree
Wear, Keith A.
Weaver, Richard L.
Webb, Spahr
Weber, Andrea
Weber, Reinhard
Weenink, David

Weisenberger, Janet
Weismer, Gary G.
Weiss, Dan
Welby, Pauline
Werner, Lynne A.
Werner, Stefan
Westwood, Evan K.
Whalen, Doug H.
Whitaker, Rodney
White, Laurence
White, Michael J.
Whitmal, Nathaniel A.
Wiegrebe, Lutz
Wightman, Fred L.
Wilcox, Paul D.
Wilhelms-Tricarico, Reiner
Wilkens, Volker
Williams, Jean-Pierre
Williams, Jon
Williams, Kevin L.
Wilmut, Mike J.
Wilson, D. Keith
Wilson, Preston S.
Winfield, Ian
Winker, Douglas F.
Withnell, Robert H.
Wolfe, Joe
Woodhouse, Jim
Worcester, Peter F.
Wouters, Jan
Wrench, Alan
Wright, Andrew B.
Wright, Beverly A.
Wright, Richard A.
Wright, William M.
Wu, Hwai-Chung
Wu, Junru
Wu, Kuangcheng
Wu, Ping
Wu, Sean F.
Wu, Ting-Wen
Wu, Tsung-Tsong
Wunenburger, Regis
Wurm, Lee
Wysocki, Lidia E.

Xiang, Ning
Xie, Bo-sun
Xu, Li
Xu, Yi
Xu, Zhen

Yang, Tsih C.
Yang, Wen
Yang, Xinmai
Yano, Takeru
Yehia, Hani C.
Yevick, David
Yin, Tom C.
Yonak, Serdar
Yoshikawa, Shigeru
Yost, William A.
Yost, William T.
Young, Eric D.
Young, Jeff
Yu, Feng
Yu, Miao
Yuan, Jiahong
Yuen, Michelle M.
Yun, Tae Sup
Yunusova, Yana

Zabolotskaya, Evgenia A.
Zagzebski, James A.
Zahorian, Stephen A.
Zahorik, Pavel
Zaitsev, Vladimir
Zampolli, Mario
Zanelli, Claudio I.
Zatorre, Robert J.
Zavorotny, Valery
Zdenek, Skvor
Zderic, Vesna
Zeddies, David
Zeitler, Berndt
Zemp, Roger J.
Zeng, Fan-Gang
Zeqiri, Bajram
Zhang, Hailan
Zhang, Honghu
Zhang, Jie
Zhang, Xiaoming
Zhang, Yu
Zhang, Zhaoyan
Zhong, Pei
Zhu, Zhemin
Zhu, Zhenya
Ziemann, Astrid
Zimmer, Walter
Zinin, Pavel
Zuckerwar, Allan J.
Zumberge, Mark
Zupan, Barbra
Zurek, Patrick

USA Meetings Calendar

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

2009

18–22 May	157th Meeting of the Acoustical Society of America, Portland, OR [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
24–28 June	5th International Middle-Ear Mechanics in Research and Otology (MEMRO), Stanford University, Stanford, CA [http://memro2009.stanford.edu].
26–30 October	158th Meeting of the Acoustical Society of America, San Antonio, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

2010

19–23 April	158th Meeting of the Acoustical Society of America, Baltimore, MD [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
15–19 November	2nd Iberoamerican Conference on Acoustics (Joint Meeting of the Acoustical Society of America, Mexican Institute of Acoustics, and Iberoamerican Federation on Acoustics), Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

ACOUSTICAL NEWS—INTERNATIONAL

W. G. Mayer
Physics Dept., Georgetown University, Washington, DC 20057

International Meetings Calendar

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

April 2009

5–9 **Noise and Vibration: Emerging Methods (NOVEM 2009)**, Oxford, UK (www.isvr.soton.ac.uk/NOVEM2009).

13–17 **2nd International Conference on Shallow Water Acoustics**, Shanghai, China ([soon]: www.apl.washington.edu).

19–24 **International Conference on Acoustics, Speech, and Signal Processing**, Taipei. R.O.C. (icassp09.com).

June 2009

2–5 **XXI Session of the Russian Acoustical Society**, Moscow, Russia (www.akin.ru/main.htm).

17–19 **3rd International Conference on Wind Turbine Noise**, Aalborg, Denmark (www.windturbinenoise2009.org)

21–25 *__13th International Conference "Speech and Computer,"__ St. Petersburg, Russia (www.specom.nw.ru).

22–26 **3rd International Conference on Underwater Acoustic Measurements: Technologies and Results**, Nafplion, Peleponnese, Greece (www.uam2009.gr).

July 2009

5–9 **16th International Congress on Sound and Vibration**, Krakow, Poland (www.icsvl6.org).

19–23 *__15th International Conference on Photoacoustics and Photothermal Phenomena__, Leuven, Belgium 9WWW.icpppl5.be).

August 2009

12–16 *__7th Triennial Conference of the European Society for Cognitive Science of Music (ESCOM2009)__, Jyväskylä, Finland (www.fyu.fi/hum/laitokset/musikki/en/escom2009).

23–28 **Inter-noise 2009**, Ottawa, Ont, Canada (www.internoise2009.com).

September 2009

6–10 **InterSpeech 2009**, Brighton, UK (www.interspeech2009.org).

9–11 **9th International Conference on Theoretical and Computational Acoustics**, Dresden, Germany (Ictca2009.com).

14–18 **5th Animal Sonar Symposium**, Kyoto, Japan (cse.fra.afrrc.go.jp/akamatsu/AnimalSonar.html).

15–17 **Autumn Meeting of the Acoustical Society of Japan**, Koriyama, Japan (www.asj.go.jp/index-en.html).

19–23 **IEEE 2009 Ultrasonics Symposium, Rome**, Italy (e-mail: pappalar@uniroma3.it).

21–23 **10th Western Pacific Acoustics Conference (WESPAC)**, Beijing, China (www.wespacx.org).

23–25 **Pacific Rim Underwater Acoustics Conference (9pruac)**, Xi'an, China (e-mail: lfh@mail.ioa.ac.cn).

23–25 **TECNIACUSTICA2010**, Cadiz, Spain (www.sea-acustica.es).

October 2009

5–7 *__International Conference on Complexity of Nonlinear Waves__, Tallinn, Estonia (www.ioc.ee/cnw09).

26–28 **Euronoise 2009**, Edinburgh, UK (www.euronoise2009.org.uk).

June 2010

9–11 *__14th Conference on Low Frequency Noise and Vibration__, Aalborg, Denmark (www.lowfrequency2010.org).

13–16 **INTERNOISE2010**, Lisbon, Portugal, (www.internoise2010.org).

August 2010

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

September 2010

26–30 **Interspeech 2010**, Makuhari, Japan (www.interspeech2010.org).

June 2011

27–1 *__Forum Acusticum 2011__, Aalborg, Denmark, (www.fa2011.org).

August 2011

27–31 **Interspeech 2011**, Florence, Italy (www.interspeech2011.org).

September 2011

4–7 **International Congress on Ultrasonics**, Gdansk, Poland.

4–7 *__Pmternoise 2011__, Osaka, Japan

June 2013

2–7 **21st International Congress on Acoustics (ICA2013)**, Montréal, Canada (www.ica20013montreal.org).

 0001-4966/2009/125(4)/2461/1/$25.00

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, PB92
California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
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*
KEVIN P. SHEPHERD, *Mail Stop 463, NASA Langley Research Center, Hampton, Virginia 23681*
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*

7,368,862

43.38.Bs ELECTROACTIVE POLYMER GENERATORS

Ronald E. Pelrine *et al.*, assignors to SRI International
6 May 2008 (Class 310/800); filed 24 January 2006

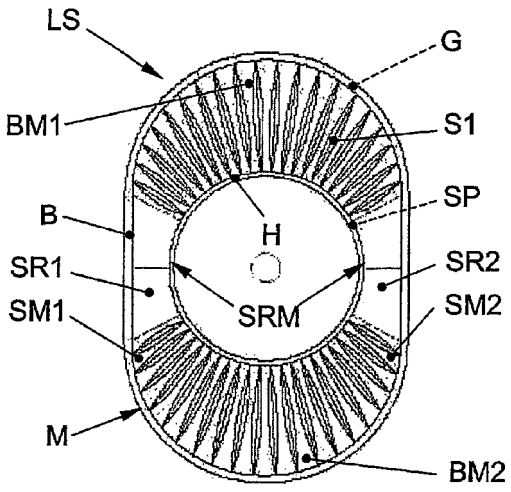
Numerous electroactuator devices are described that use electroactive polymers including an artificial muscle, a flex-tensional actuator, and various bending beams. Methods of enhancing the mechanical forces created in the electromotive polymer are claimed.—AJC

7,416,047

43.38.Ja DIAPHRAGM FOR A LOUDSPEAKER WITH A MOVING COIL

Ewald Frasl, A-2362 Biedermannsdorf and Helmut Wasinger, A-2371 Hinterbruehl, both of Austria
26 August 2008 (Class 181/173); filed in the European Patent Office 29 April 2004

Stress reduction areas **SR1**, **SR2** of diaphragm **M** do not have V-shaped corrugations **S1** as do wide diaphragm portions **BM1**, **BM2**. This enhancement over prior art, in which corrugations **S1** cover the whole of the diaphragm area, is said to reduce the stiffness in areas **SR1**, **SR2**—which are stiff due to their small area between retaining part **H** and fastening part **B**—and so equalizes the stiffness of the narrow and wide areas, reducing the "uneven load on the moving coil **SP**."—NAS

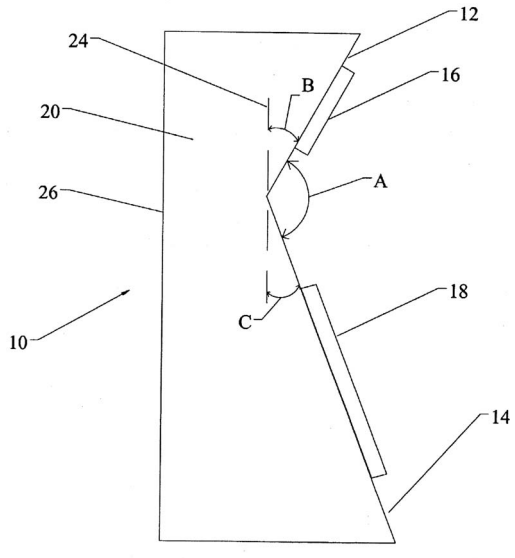


7,418,104

43.38.Ja SPEAKER SYSTEM

Robert G. Spinale, Middleton, New Hampshire
26 August 2008 (Class 381/87); filed 6 August 2002

Loudspeaker **10** is comprised of four speakers—one full range speaker **16** in upper baffle **12** and woofer **18**, a midrange speaker (not shown in this figure), and a tweeter type speaker (not shown) mounted in lower baffle **14**. The novel part is that upper **12** and lower **14** baffles are canted relative to one another so that the angle **A** between them is between 120 and 140 degrees (which can change depending on the embodiment).—NAS



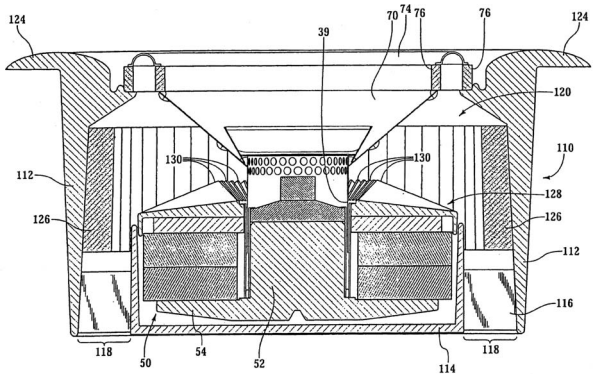
 0001-4966/2009/125(4)/2462/12/$25.00

7,443,995

43.38.Ja ACOUSTIC LOUDSPEAKER

Burton Babb and Alan Babb, assignors to Babb Laboratories
28 October 2008 (Class 381/396); filed 17 February 2005

Perhaps the title of this patent is intended as a kind of Zen exercise, inviting the reader to contemplate the attributes of a nonacoustic loudspeaker. The inventor has patented several earlier loudspeaker designs, most of which utilize a sliding bearing rather than a spider to center the voice coil. The voice coil of this latest variant is a single layer of round or square wire with the windings spaced apart, the width of the gaps being roughly equal to



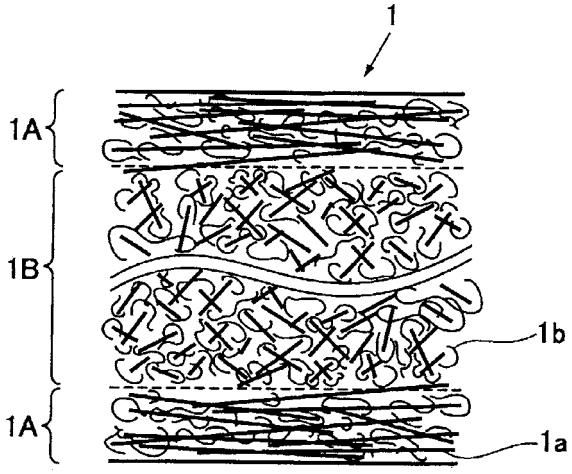
the diameter of the wire. "Thus, contrary to traditional teachings for an efficient loudspeaker 'motor' or transducer, such a coil has better magnetic coupling. With better magnetic coupling and less resistance due to the relatively shorter length of wire used, it is more efficient at converting electrical to acoustic energy and has less mass for a given length than conventional prior art coils." Well, the assertion of better magnetic coupling is certainly easy to test, and one can only hope that the inventor will eventually get around to testing it.—GLA

7,443,998

43.38.Ja SPEAKER DIAPHRAGM AND MANUFACTURING METHOD THEREOF

Satoshi Hachiya, assignor to Pioneer Corporation
28 October 2008 (Class 381/426); filed in Japan 4 June 2002

Over the years, Pioneer has patented a number of new materials and manufacturing techniques for loudspeaker diaphragms. In this patent, fibers from 3 to 5 mm in length are mixed in a matrix resin. The illustration is



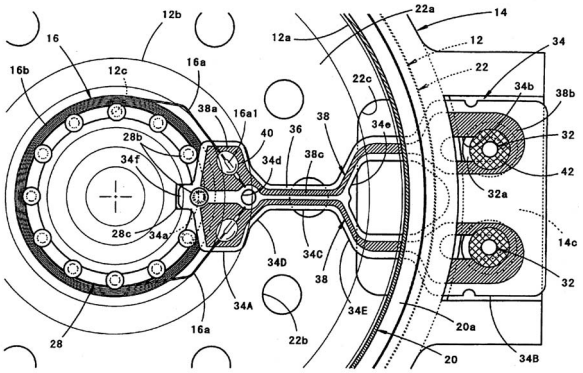
intended to show that the fibers are oriented radially in the surface layers and randomly in the central layer of the diaphragm.—GLA

7,447,327

43.38.Ja FLEXIBLE PCB VOICE COIL CONNECTOR

Hajime Kitamura *et al.*, assignors to Star Micronics Company, Limited
4 November 2008 (Class 381/410); filed in Japan 11 May 2004

Element **34** is a small, lightweight, flexible printed circuit board that provides electrical connections between terminals **32** and voice coil leads **16a** of a miniature loudspeaker. The central portion **34c** is made narrower



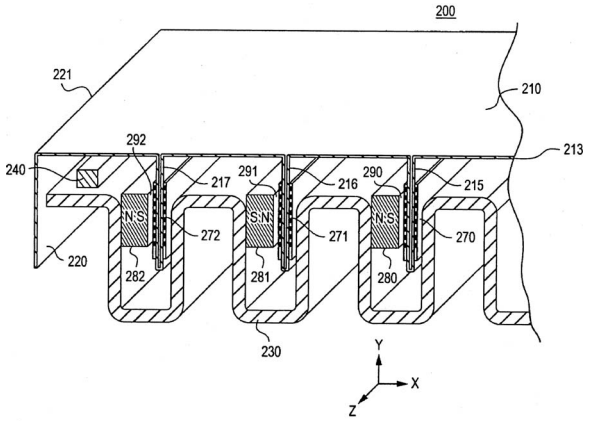
than the ends to minimize stiffness and localized stresses. In practice, the assembly shown is quite small—section **34c** is only about 2 mm wide and less than 0.5 mm thick.—GLA

7,450,729

43.38.Ja LOW-PROFILE TRANSDUCER

An Duc Nguyen and Charles M. Sprinkle, assignors to Harman International Industries, Incorporated
11 November 2008 (Class 381/152); filed 9 April 2004

Elongated rectangular voice coils have been a feature of some planar loudspeaker designs for more than 40 years, and at least half of the claims of



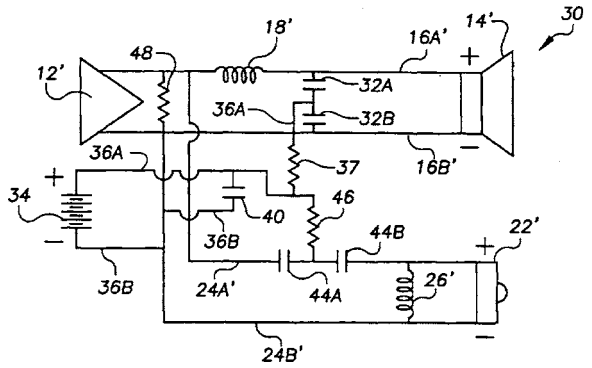
this patent appear to describe prior art. Nonetheless, the magnetic circuit shown is clever, simple, and practical.—GLA

7,443,990

43.38.Lc VOLTAGE BIASED CAPACITOR CIRCUIT FOR A LOUDSPEAKER

Daniel A. Chattin, Harwinton, Connecticut
28 October 2008 (Class 381/111); filed 1 November 2004

Some high-end loudspeaker systems include DC bias for crossover network capacitors. This is usually done to stabilize the long-term perfor-



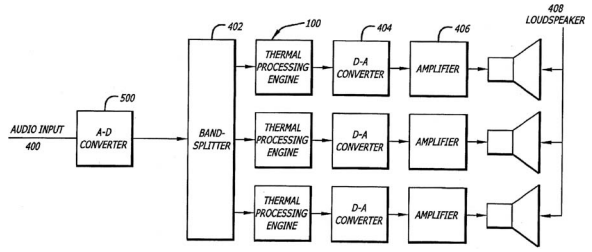
mance of electrolytic capacitors, but is sometimes said to improve the performance of non-polarized capacitors as well. This patent discloses a biasing method using series pairs of capacitors **32**, **44** biased through resistors **37** and **46**. Note that the diagram is drawn incorrectly—resistor **37** connects directly to **36A** and not to **16B**.—GLA

7,447,318

43.38.Lc SYSTEM FOR USING DIGITAL SIGNAL PROCESSING TO COMPENSATE FOR POWER COMPRESSION OF LOUDSPEAKERS

Douglas J. Button and Paul Robert Williams, assignors to Harman International Industries, Incorporated
4 November 2008 (Class 381/55); filed 7 September 2001

Multi-band loudspeaker systems, especially high-power systems, are commonly used in conjunction with a signal processor designed for the particular components employed. The power of digital processing allows the processor to include all sorts of sophisticated functions such as frequency division, equalization, delay alignment, and overload protection. To permit the greatest possible acoustic output without damage, overload protection



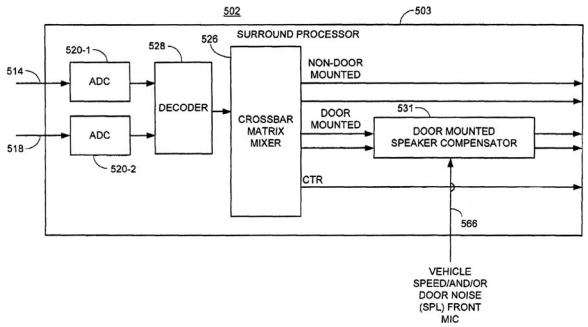
must sense or estimate cone excursion and voice coil temperature. This patent teaches that a computerized thermal model can accurately predict voice coil temperature without the need for additional sensing devices. The predicted value can then be used in real time to protect against overload and also to compensate for frequency response shifts caused by the coil's resistance vs temperature characteristic. The patent text also suggests that the proposed method can be applied to other devices such as motors and transformers. Such applications appear to be in the public domain, however, since they are not covered by any of the 36 patent claims.—GLA

7,451,006

43.38.Lc SOUND PROCESSING SYSTEM USING DISTORTION LIMITING TECHNIQUES

Bradley F. Eid and William Neal House, assignor to Harman International Industries, Incorporated
11 November 2008 (Class 700/94); filed July 2002

This is one of a three-patent package having to do with stereo signal processing in automotive sound systems. Presumably the other two patents are still pending. Looking at any one patent, this one-into-three ploy results in numerous illustrations and long stretches of descriptive text that have little or nothing to do with the patent claims. In the case at hand, the claims seem to cover the concept of dynamically equalizing various channels of a matrixed multi-channel audio system in response to the volume setting and background noise.—GLA

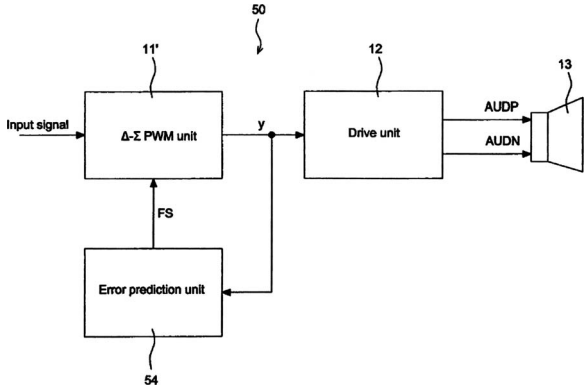


7,439,440

43.38.Md AUDIO PLAYER USING SIGMA-DELTA PULSE-WIDTH MODULATION

Han-Huah Hsu, assignor to Sunplus Technology Company Limited
21 October 2008 (Class 84/600); filed in Taiwan 11 October 2004

The use of pulse-width modulation (Class D) amplifiers in audio playback devices has become commonplace. However, using feedback to reduce



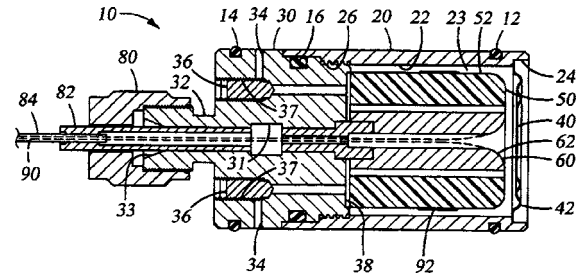
distortion in Class D amplifiers is difficult and costly. This patent asserts that distortion produced by drive unit **12** under loading is predictable, and can therefore be reduced through predistortion rather than feedback. Error prediction unit **54** may use a simple lookup table to generate the required correction. How the process can be accomplished in the digital domain is thoroughly explained in the patent.—GLA

7,369,716

43.38.Pf HIGH PRESSURE AND HIGH TEMPERATURE ACOUSTIC SENSOR

Arne Berg *et al.*, assignors to Weatherford/Lamb, Incorporated
6 May 2008 (Class 385/12); filed 9 March 2004

Borehole hydrophone package **10** having a fiber optics sensor capable of operating in well down-hole high-pressure and high-temperature conditions is claimed. Environmental pressure and sound are admitted through



diaphragm **40** into chamber **22** containing fluid **23** and Teflon mandrel **50** having optical fiber spiral sensor **92** wound on its outer surface. Fiber ends **90** exit through the center of the mandrel **50** and through supporting cable **84**.—AJC

6,906,632

43.38.Si VEHICULAR SOUND-PROCESSING SYSTEM INCORPORATING AN INTERIOR MIRROR USER-INTERACTIION SITE FOR A RESTRICTED-RANGE WIRELESS COMMUNICATION SYSTEM

Jonathan E. DeLine *et al.*, assignors to Donnelly Corporation
14 June 2005 (Class 340/815.4); filed 8 July 2002

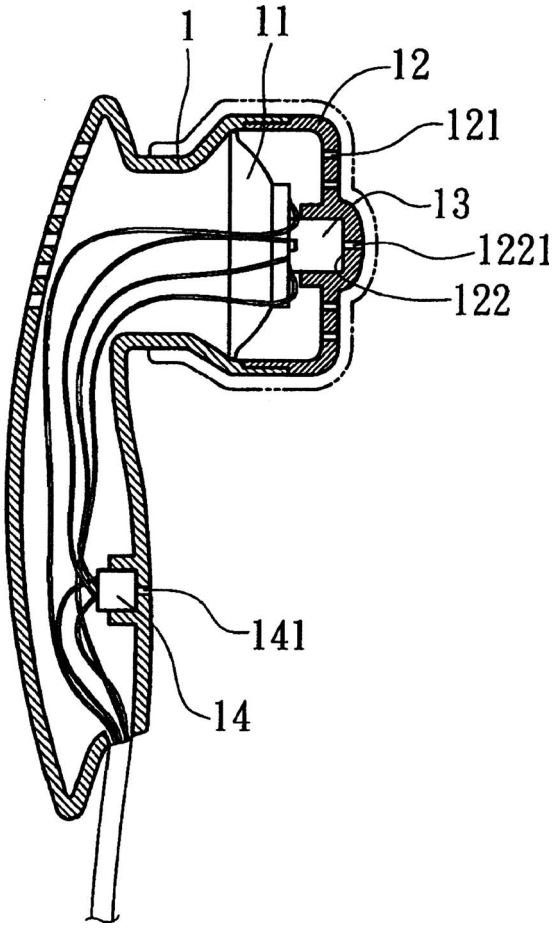
This huge patent, with 517 claims, is primarily based on a wireless control system that is integral to the rear view mirror assembly in an automobile. A voice recognition system is used to control accessories such as cell phones, internet connections, and in-vehicle systems.—KPS

7,447,308

43.38.Si LOW-NOISE TRANSMITTING RECEIVING EARSET

Jin-Chou Tsai, Taipei City, Taiwan
4 November 2008 (Class 379/430); filed 26 August 2005

This noise-cancelling headset is an improved version of an earlier design by the same inventor. Sound from the rear side of loudspeaker **11** reaches the user's ear through holes **121**. Microphone **13** picks up "resonant sound waves from the external auditory canal of the user's ear," and second microphone **14** picks up the user's voice. It appears that the two microphones are connected in parallel. The patent gives no additional clues as to how this arrangement is intended to operate.—GLA

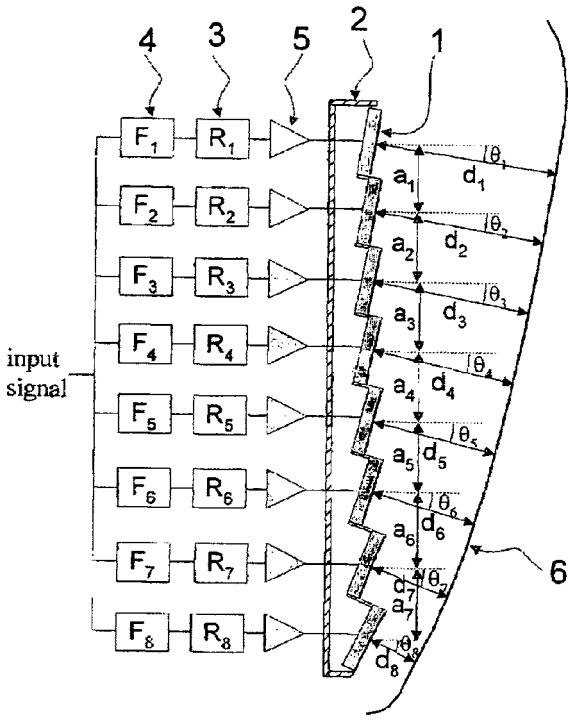


7,426,278

43.38.Tj SOUND DEVICE PROVIDED WITH A GEOMETRIC AND ELECTRONIC RADIATION CONTROL

Xavier Meynial, assignor to Active Audio
16 September 2008 (Class 381/82); filed in France 25 March 2004

Elements **1** of column loudspeaker **2** are angled as shown. Elements are connected to amplifiers **5** which are connected to delays **3** and filters **4**. Elements can be planar type radiators, a combination of planar type radiators, cone type radiators, or even transducer/horn assemblies. "The advantage of the system that is the object of this invention is to combine the advantages of a geometrical network with those of an electronically controlled network…" which allows for control of the main and secondary lobes. The patent describes a device designed by the reviewer in the late 1980s, and he was probably not the first.—NAS

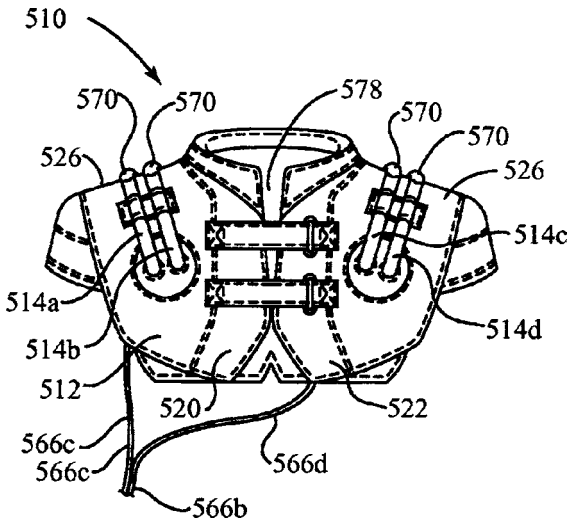


7,440,581

43.38.Vk BACKPACK WITH INTEGRATED SPEAKERS

David Wiener, assignor to Soundtube Entertainment, Incorporated
21 October 2008 (Class 381/388); filed 24 April 2006

Four loudspeakers are attached to backpack **510**: two in front and two on the back. Pairs of loudspeakers are loaded by flexible ducts **514** which



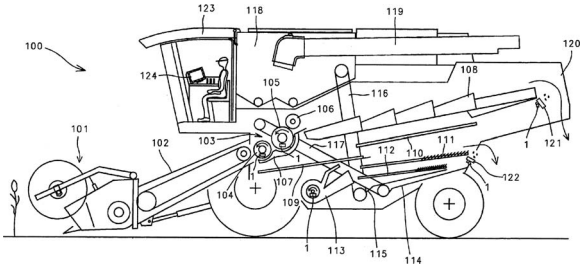
may be circular or rectangular in cross-section, and may include vents or perforations. The ducts are configured with stiffeners or corrugations to prevent collapse during flexing yet not restrict the wearer's movements. Performance goals include enhanced low-frequency response and a physical sensation to the body of the wearer.—GLA

7,415,365

43.40.Le STRUCTURE-BORNE SOUND SENSOR UNIT

Eckehard Jeppe, assignor to CLAAS Selbstfahrende Erntemaschinen GmbH
19 August 2008 (Class 702/56); filed in Germany 30 March 2006

A means of determining the amount of waste grain that exits from straw-walker **108** and from upper sieve **11** in a combine harvester is disclosed that uses structure borne sound sensors **1** that are attached to pulse detectors **121**, **122**. Sensors **1** are multi-purpose units which contain signal conditioning circuitry and so can be used in many situations, another of which can be the chopper drum in a forage harvester. The output from sensors **1** are connected to processing unit **124**. The patent provides a concise description of how these harvesters work.—NAS



7,438,530

43.40.Tm METHOD OF ASSEMBLING ONE-PIECE BLADED DISKS, AND A DEVICE FOR DAMPING VIBRATION OF THE BLADES OF SUCH DISKS

Jean-Pierre Ferte *et al.*, assignors to SNECMA
21 October 2008 (Class 416/198 A); filed in France 17 April 2004

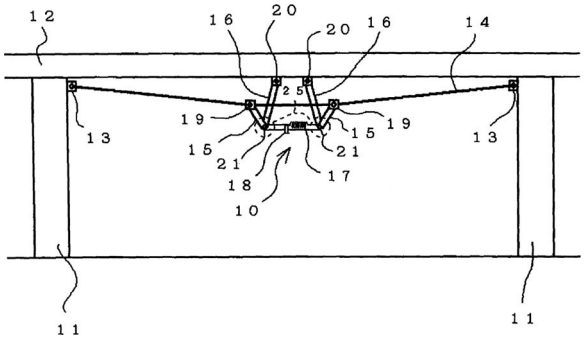
In friction-welding of blades onto a disk, the blades may suffer damage from excessive vibrations. To overcome this problem the blades are damped during the assembly process by clamping them in an annular assembly in which elastomeric elements are pressed against the blade surfaces.—EEU

7,441,376

43.40.Tm BASE ISOLATION DEVICE FOR STRUCTURE

Shinji Ishimaru *et al.*, assignors to Nihon University, School Juridical Person
28 October 2008 (Class 52/167.6); filed in Japan 26 December 2001

The device described in this patent is intended to suppress out-of-plane vibrations of plate-like structural components, such as slabs of elevated



roads or bridges, or roof structures, or glass curtain walls. Cables or other tension members **14** are attached to columns or bridge supports **11** at points **13** underneath the slab **12**. A linkage arrangement is pin-connected to the slab at **20** and to the tension member at **19**, so that out-of-plane deflection of the slab induces axial movement in the damper **18**. The spring **17** is intended to keep the assembly in tension, so that the damper can act during both upward and downward deflection of the slab.—EEU

7,445,685

43.40.Tm ARTICLE HAVING A VIBRATION DAMPING COATING AND A METHOD OF APPLYING A VIBRATION DAMPING COATING TO AN ARTICLE

Martin J. Deakin *et al.*, assignors to Rolls-Royce PLC
4 November 2008 (Class 156/212); filed in United Kingdom 23 March 2004

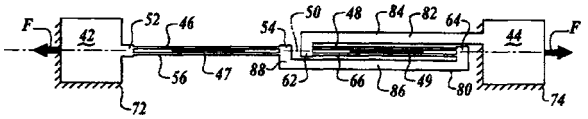
This patent pertains to vibration-damping coatings for blades or vanes of fans, compressors, turbines, and the like. A damping coating, such as a polymer-impregnated ceramic, is deposited on an erosion-resistant material, such as stainless steel, by plasma spraying. The coating may be applied through a mesh mask, which is then removed, leaving the coating in separated segments. The assembly is bonded adhesively to the article to be damped, with the damping coating in contact with the article—so that there results a constrained-layer damping arrangement.—EEU

7,444,883

43.40.Yq VIBRATING BEAM FORCE TRANSDUCER

Mitchell J. Novack, assignor to Honeywell International Incorporated
4 November 2008 (Class 73/826); filed 22 December 2005

Vibrating beam force transducers may be used as force sensors, for example in accelerometers and pressure sensors. It is often advantageous to employ two such transducers in a sensor, with one subjected to tension and the other to compression, in order to cancel common mode errors, such as may occur due to temperature sensitivity and some nonlinearities. This method of error canceling is effective only to the extent that the two force sensing elements experience the same force loading; thus, this patent describes configurations that provide such equal loading and that can be formed on a crystalline substrate. In the illustrative embodiment shown in the figure an arm **82** is connected to the force application element **44**. Sensing element **48** is connected between that arm and a frame **86**. Sensing



element **46** is connected between that frame and the force application element **42**. The sensing elements are aligned so that their axes **47** and **49** are along the axis **50** of the force transducer.—EEU

7,448,268

43.40.Yq MECHANICAL DECOUPLING DEVICE FOR MONOLITHIC DIFFERENTIAL VIBRATING SENSOR

Olivier Le Traon *et al.*, assignors to Onera
11 November 2008 (Class 73/514.29); filed in France 5 May 2004

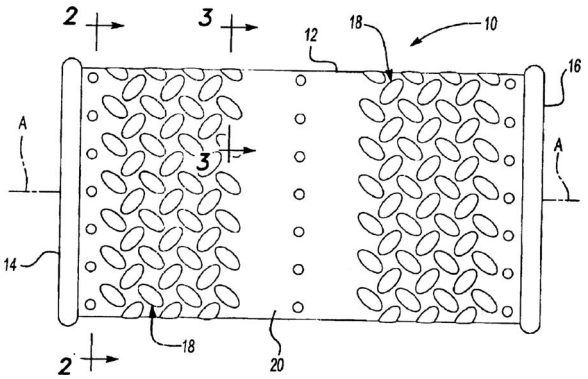
Monolithic differential vibrating sensors consist in essence of two substantially identical vibrating members attached to the same support. Each of the vibrating members is an accelerometric transducer consisting of a proof mass and a resonator whose resonant frequency varies as a function of the applied acceleration. The difference between these two frequencies, which serves as a measure of the applied acceleration, is relatively insensitive to such parasitic disturbances as temperature variations. In order to minimize the undesirable mechanical coupling between the two transducer elements, the support frame for the transducers is designed to have nodes (at the system's working frequency) at the points where the transducers are connected to the frame.—EEU

7,892,855

43.50.Gf EMBOSSED SHELL FOR SPUN MUFFLERS

John Lewis Warmenhoven, assignor to Tenneco Automotive Operating Company Incorporated
17 May 2005 (Class 181/282); filed 25 February 2003

An exhaust muffler having a non-circular (e.g. elliptical) cross-section may have resonance frequencies that coincide with engine frequencies. Embossments (as shown in the figure) are stamped into the surface to stiffen it and thereby raise the resonance frequencies.—KPS



6,905,091

43.50.Gf SYSTEM AND METHOD FOR CONTROLLING THE ACOUSTIC SIGNATURE OF A DEVICE

Barry L. Berson *et al.*, assignors to Supersonic Aerospace International, LLC
14 June 2005 (Class 244/1 N); filed 14 July 2003

Despite the broad, generic title of this patent, it really concerns the control of levels of sonic booms heard on the ground from a supersonic aircraft. An electronic system receives information concerning current flight conditions and determines the resultant boom level at the ground surface. This is compared to a desired level and cues are presented to the pilot regarding corrective actions that can be taken to reduce or maintain the level at the desired value. Options to limit maneuvers and to automatically adjust operating parameters can be enabled. Options to display past, current, and future levels can be selected.—KPS

6,911,158

43.50.Gf SOUND ABSORBING MATERIAL AND ITS MANUFACTURING METHOD

Tsutomu Oishi *et al.*, assignors to Ishikawajima-Harima Heavy Industries Company, Limited
28 June 2005 (Class 252/62); filed in Japan 29 November 2000

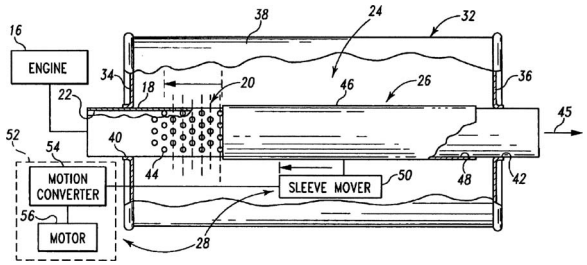
A porous sound absorbing material suitable for hostile environments, such as found in the exhaust duct of a jet engine, is described. A ceramic foam material having 75–80% porosity and a range of cell sizes is designed for broadband absorption. The manufacturing process is composed of two steps: in the first, the raw material containing ceramic grains and a void-forming material (carbon) is formed into a prescribed shape and, in the second, the carbon is removed. Details are provided concerning material selection and performance.—KPS

6,915,876

43.50.Gf EXHAUST PROCESSOR WITH VARIABLE TUNING SYSTEM

Mehmet S. Ciray, assignor to Arvin Technologies, Incorporated
12 July 2005 (Class 181/219); filed 1 December 2003

An exhaust muffler for an automobile includes an adjustable Helmholtz resonator that allows the tuning frequency to match the engine conditions. The tuning is accomplished by means of a movable sleeve that effectively adjusts the throat of the resonator.—KPS

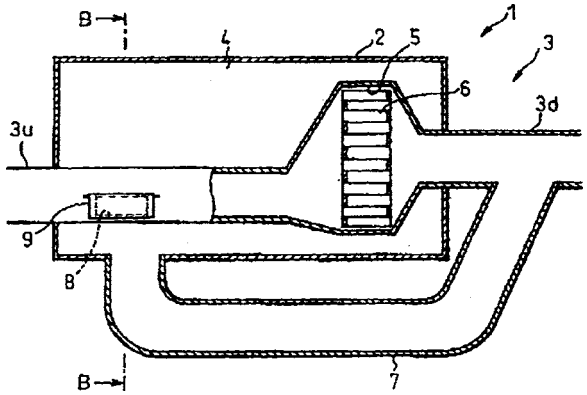


6,918,463

43.50.Gf MUFFLER FOR ENGINE

Masanori Takahashi *et al.*, assignors to Toyota Jidosha Kabushiki Kaisha
19 July 2005 (Class 181/237); filed in Japan 21 May 2002

An engine exhaust muffler containing a particulate filter, **6**, also includes a control valve, **9**. At low engine speeds the valve is closed and all



exhaust gases flow through the filter. At higher speeds the valve opens and a portion of the flow is diverted, thus keeping backpressure low.—KPS

6,920,958

43.50.Gf ANNULAR ACOUSTIC PANEL

Geoffrey E. Harrison, assignor to The Boeing Company
26 July 2005 (Class 181/214); filed 17 October 2003

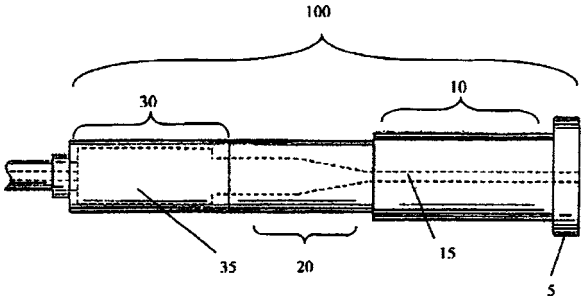
Acoustic absorption treatment in a turbofan engine nacelle is generally installed in sections, resulting in seams and untreated areas. The invention describes a monolithic one-piece design that eliminates seams and provides improved fan noise reduction.—KPS

6,923,292

43.50.Gf ATTACHMENT FOR RECOIL, NOISE, BLAST AND FLASH SUPPRESSION OF THERMODYNAMIC JETTING DEVICES SUCH AS FIREARMS, HIGH PRESSURE EXHAUST MECHANISMS AND OTHER HEAT ENGINE DEVICES, WHICH PRODUCE SUCH JETTING EXHAUST ACTION AS A RESULT OF THEIR FUNCTION

Robert James Woods, Phoenix, Arizona and JeanMarie Evangeline Serrano, Danville, California
2 August 2005 (Class 181/223); filed 15 April 2003

A silencer for a gun is attached to the barrel and consists of "a primary chaotic expansion chamber, **30**, the secondary convergent compression chamber, **20**, and, lastly, the magnetic diversion chamber assembly, **10**." The



latter consists of rare-earth magnetic toroids that "redirect, helically confine and delay, cool, and extract energy from the gaseous exhaust column." No acoustic data are provided.—KPS

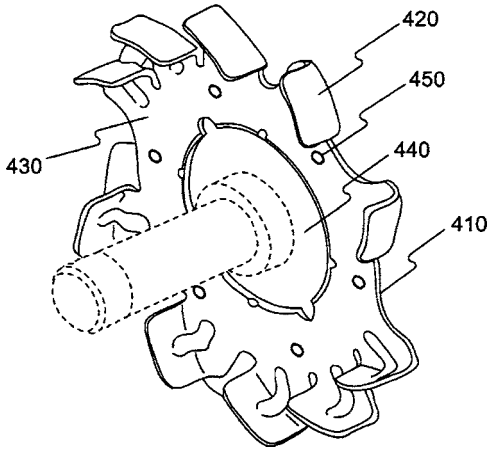
7,365,471

43.50.Gf INTERNAL COOLING FAN WITH A NON-REPEATING BLADE CONFIGURATION

Alex Creviston *et al.*, assignors to Remy Incorporated
29 April 2008 (Class 310/263); filed 28 October 2005

The pitch angle of blades **420** of the internal cooling fan **220** that is attached to the rotating rotor (not shown) of a motor, a generator, or an

220



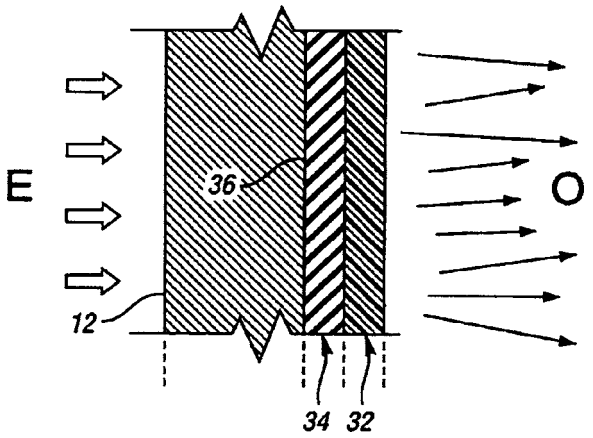
alternator is varied from one blade **420** to the next to reduce the intensity of the blade passage frequency noise.—AJC

7,367,304

43.50.Gf APPARATUS AND METHOD FOR FORCED RESPONSE ACOUSTIC ISOLATION ENCLOSURE IN CAST ALUMINUM OIL PAN

David T. Hanner *et al.*, assignors to GM Global Technology Operations, Incorporated
6 May 2008 (Class 123/195 C); filed 2 December 2005

Cladding **E-O** for the exposed aluminum oil pan bottom of a reciprocating internal combustion engine that reduces the noise emitted therefrom is claimed. Foam rubber isolator **34** is bonded to oil pan **12** outer surface **36**.



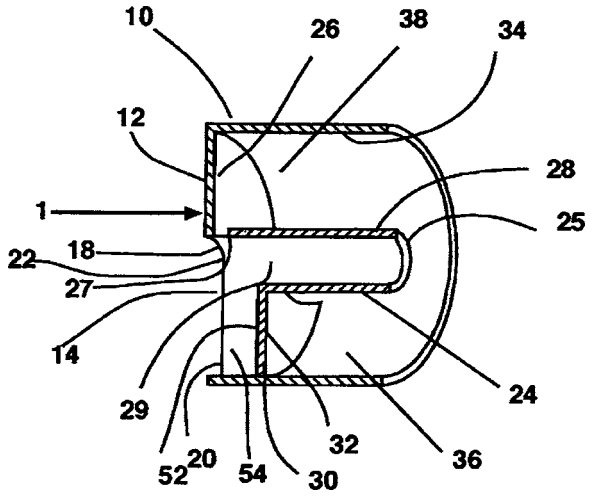
Urethane sound barrier **32** is affixed over all pan surfaces including surfaces that may not be covered by isolator **34**.—AJC

7,412,917

43.50.Gf SOUND SUPPRESSOR SILENCER BAFFLE

George Vais, Boisè, Idaho
19 August 2008 (Class 89/14.4); filed 13 December 2005

Suppressor baffle segment **1** has bulkheads **12** and **32** offset from one another, and central core defining tube **24** contains a longitudinal slot **28** that extends along the top portion of the central core **24** to intersection **37** of rear wall **12**. This provides a relief path for the propellant gasses while also, it is claimed, reducing the effect of gas flow on the projectile. A complete sup-



pressor is comprised of several baffle segments, a means of attachment to the muzzle of the firearm, and an end cap.—NAS

7,448,468

43.50.Gf ACOUSTICALLY EFFECTIVE WHEEL HOUSE COVERING FOR VEHICLES

Hans Rudolf Czerny *et al.*, assignors to Carcoustics Tech Center GmbH
11 November 2008 (Class 181/210); filed in Germany 15 July 2003

A liner to be placed in the wheel housings of automotive vehicles in order to reduce the transmission of water splash noise into the passenger compartment and to absorb tire/road noise consists of three or more air-permeable sandwich layers. The layers, typically of textile materials, have graded permeabilities, thicknesses, and acoustic impedances.—EEU

6,895,290

43.55.Dt LIMITING THE DAMAGING EFFECTS OF LOUD MUSIC FROM AUDIO SYSTEMS, PARTICULARLY FROM AUTOMOBILE AUDIO SYSTEMS

Faisal M. Awada *et al.*, assignors to International Business Machines Corporation
17 May 2005 (Class 700/94); filed 11 July 2002

A scheme that limits the loudness of music played over an automobile sound system is described. Music having predominantly low frequency content results in limiting of the amplifier gain since this is the kind of music most likely to be played at high levels. A nice idea, but not likely to catch on!—KPS

6,896,219

43.55.Dt PROCESS AND APPARATUS FOR NOISE REDUCTION IN MULTI-ENGINE PROPELLER-DRIVEN AIRCRAFT

Ingo Udo Borchers *et al.*, assignors to Dornier GmbH
24 May 2005 (Class 244/1 N); filed in Germany 19 March 2002

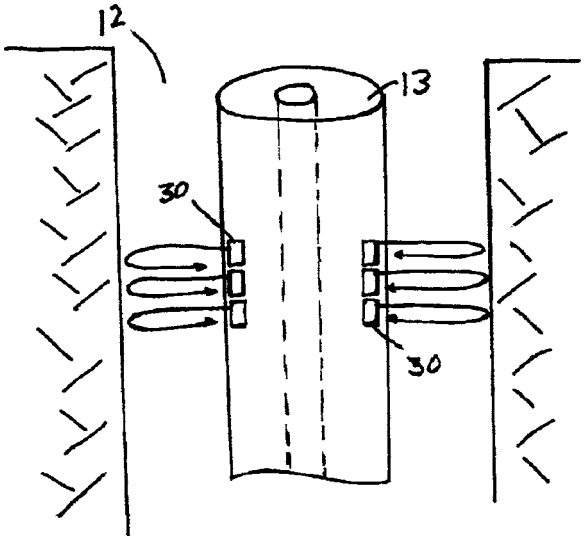
A system to reduce cabin noise of propeller aircraft is based on destructive interference of sound fields from two propellers. Unlike previous schemes based on synchrophasing propellers on opposite sides of the fuselage, this one proposes to use interference between sound fields from adjacent propellers on the same wing. It is thus necessary to compensate for the difference in distance from each propeller to the fuselage sidewall. Various methods are proposed to alter the acoustic fields of the two adjacent propellers, such as different numbers of blades, different rotation speeds, blade geometry, and relative upstream location. No data are presented to indicate that any such methods might work.—KPS

7,364,007

43.60.Qv INTEGRATED ACOUSTIC TRANSDUCER ASSEMBLY

Fernando Garcia-Osuna *et al.*, assignors to Schlumberger Technology Corporation
29 April 2008 (Class 181/108); filed 30 November 2004

An advanced acoustical borehole sensor package is claimed that combines a miniaturized electronics processor with an array of receiver transducer elements that can be variously connected to effect monopole, dipole, and other multipole sound sensor configurations. Shown is a pulse-echo



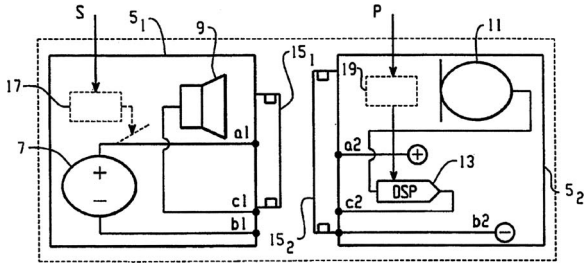
version of the invention where down-hole package **13** has several axially and azimuthally distributed receiver elements **30** which can be switched to perform directional pulse-echo measurements in various directions.—AJC

7,440,579

43.66.Ts HEARING DEVICE HAVING TWO MODULES FOR ASSEMBLING/DISASSEMBLING DEVICE

Andy Vonlanthen, assignor to Phonak AG
21 October 2008 (Class 381/322); filed 21 February 2007

The motivation here is standardization while forming a family of in-the-ear or over-the-ear hearing devices having different audio output power levels. An interchangeable first module, which has an on-off control and is



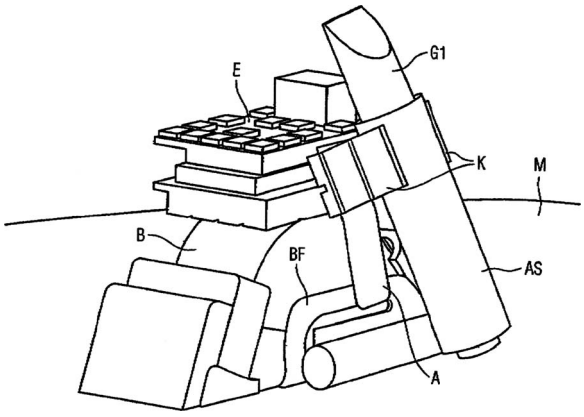
transmission-power-level dependent, houses different batteries including rechargeable types. A standardized second module that is relatively independent of power level, which is interconnected removably to the first module, contains the microphone and digital signal processor.—DAP

7,443,991

43.66.Ts ITE HEARING AID FOR BINAURAL HEARING ASSISTANCE

Torsten Niederdränk *et al.*, assignors to Siemens Audiologische Technik GmbH
28 October 2008 (Class 381/315); filed in Germany 31 March 2004

Antennae are positioned in the left and right hearing aids in a binaural fitting so they will align for good ear-to-ear wireless communication. This is achieved in in-the-ear as well as completely-in-the-canal models by mounting the antennae on a battery contact or on a programming connector at a 10 to 40 degree angle with respect to a perpendicular to the hearing aid faceplates.—DAP

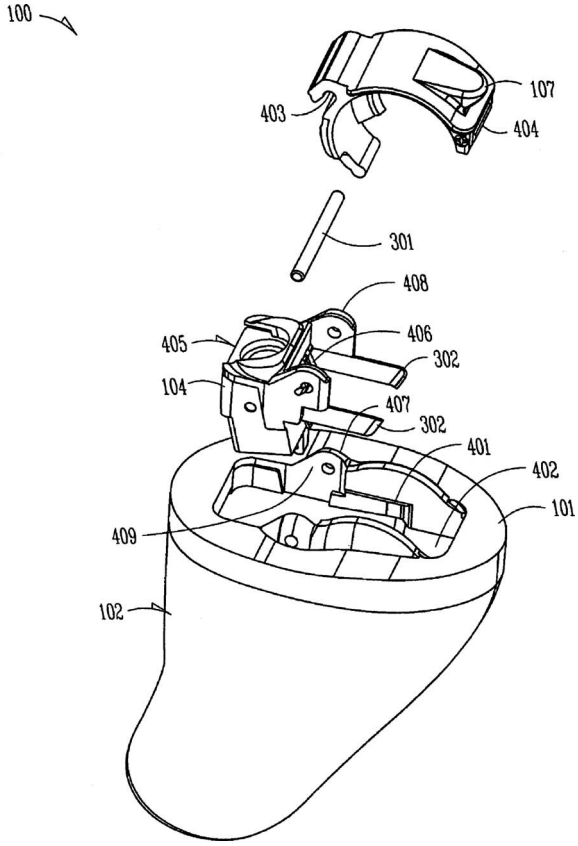


7,443,992

43.66.Ts METHOD AND APPARATUS FOR MODULAR HEARING AID

Brian Fideler, assignor to Starkey Laboratories, Incorporated
28 October 2008 (Class 381/328); filed 15 April 2004

Presented here is a method to assemble a standard-sized microphone housing that fits through an access port in the hearing aid that may or may not have a separate faceplate. The battery contacts, programming terminals, telecoil, signal processing electronics, and receiver interface are in close proximity to the microphone housing and are interconnected electrically with a flexible substrate. A removable fastening pin may be used to retain the microphone housing and adjoining components in place.—DAP

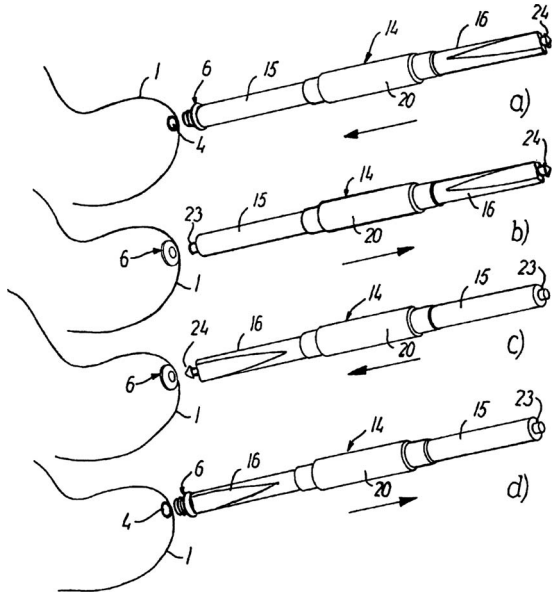


7,443,993

43.66.Ts TOOL FOR INSERTION AND REMOVAL OF A HEARING AID EAR WAX GUARD AND A METHOD FOR ITS USE

Finn Gunnersen and Jan Tøpholm, assignors to Widex A/S
28 October 2008 (Class 381/329); filed in Denmark 10 July 1998

A rod-shaped insertion-removal tool has a stopping abutment and a harpoon-shaped catch element to engage an inside wall in the mouth of the wax guard. When the tool is removed from the hearing aid without rotation, the wax guard comes out with it.—DAP



7,444,877

43.66.Ts OPTICAL WAVEGUIDE VIBRATION SENSOR FOR USE IN HEARING AID

Guann-Pyng Li *et al.*, assignors to The Regents of the University of California
4 November 2008 (Class 73/651); filed 20 August 2003

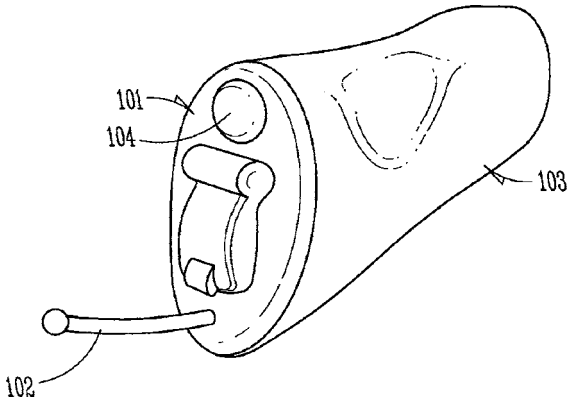
The intensity of an LED light source array output is modulated by a bank of polymeric optical waveguide cantilever resonators according to acoustic or seismic vibrations that are sensed. The resonators, which act as a quasi-Fourier analyzer, are oriented perpendicular to the direction of incoming vibration waves. A photodiode light detector array converts the resulting modulated light into electrical signals which are amplified and sent to a loudspeaker or the electrode array of a cochlear implant. The entire device is manufactured with a combination of traditional semiconductor processes and polymer microfabrication techniques.—DAP

7,446,720

43.66.Ts METHOD AND APPARATUS FOR A WIRELESS HEARING AID ANTENNA

Thomas A. Victorian and James Newton, assignors to Starkey Laboratories, Incorporated
4 November 2008 (Class 343/718); filed 19 February 2007

The insertion-removal handle for a small custom hearing aid incorporates at least a part of a coiled antenna for wireless communication.—DAP

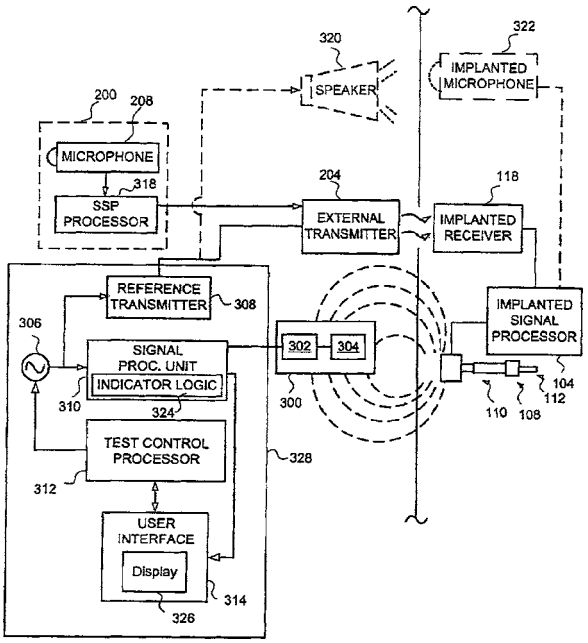


7,447,319

43.66.Ts METHOD AND SYSTEM FOR EXTERNAL ASSESSMENT OF HEARING AIDS THAT INCLUDE IMPLANTED ACTUATORS

Douglas Alan Miller and Scott Allan Miller III, assignors to Otologics, LLC
4 November 2008 (Class 381/60); filed 13 April 2006

An external test device estimates the electrical signal passing through an implanted actuator by measuring the magnetic field generated by the



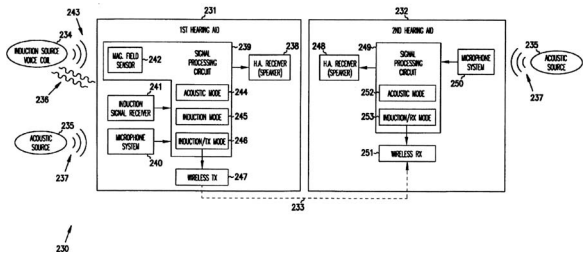
implanted transducer. This signal is an estimate of the electrical impedance of the actuator, which is indicative of the mechanical connection between the actuator and middle ear. Two inductive pickup coils in the test device, which are coaxially positioned in opposite polarity to reduce ambient electromagnetic interference, are used also to signal when the test device is properly aligned relative to the implanted actuator.—DAP

7,447,325

43.66.Ts SYSTEM AND METHOD FOR SELECTIVELY COUPLING HEARING AIDS TO ELECTROMAGNETIC SIGNALS

Mark A. Bren *et al.*, assignors to Micro Ear Technology, Incorporated
4 November 2008 (Class 381/331); filed 12 September 2002

A hearing aid normally processes acoustic signals through a microphone unless a magnetic field sensor detects the presence of a nearby inductive field from a telephone handset. If an inductive field is detected, input to the signal processing circuitry switches automatically away from the microphone to the telephone signal, and the telephone signal is also transmitted wirelessly to a second hearing aid, allowing the telephone input to be heard diotically in both ears. When the telephone handset is no longer adjacent to the ear, the input is switched automatically back to the microphone.—DAP

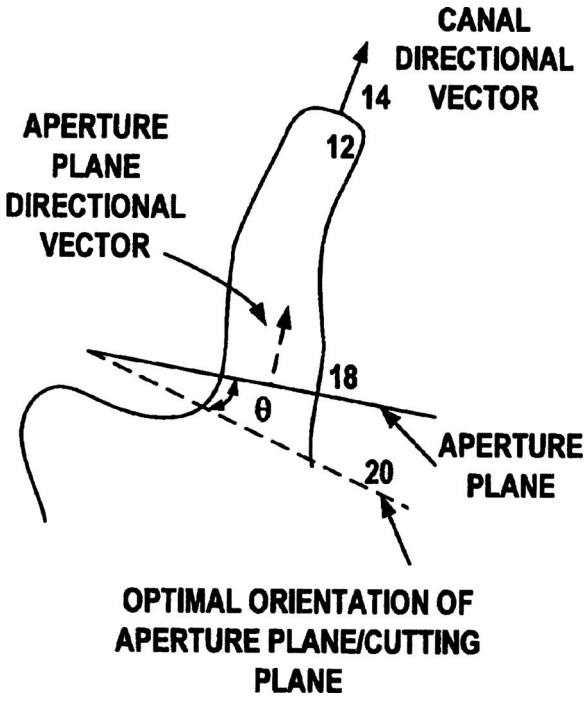


7,447,556

43.66.Ts SYSTEM COMPRISING AN AUTOMATED TOOL AND APPERTAINING METHOD FOR HEARING AID DESIGN

Fred McBagonluri *et al.*, assignors to Siemens Audiologische Technik GmbH
4 November 2008 (Class 700/98); filed 3 February 2006

A digitized geometric model is created of each of the two scanned ear impressions used for a binaural hearing aid fitting. The models are electronically detailed after a manufacturing software tool determines each ear canal direction and the locations of the first and second bends and the bony part of the ear canals. A cutting plane through the aperture of the impression is



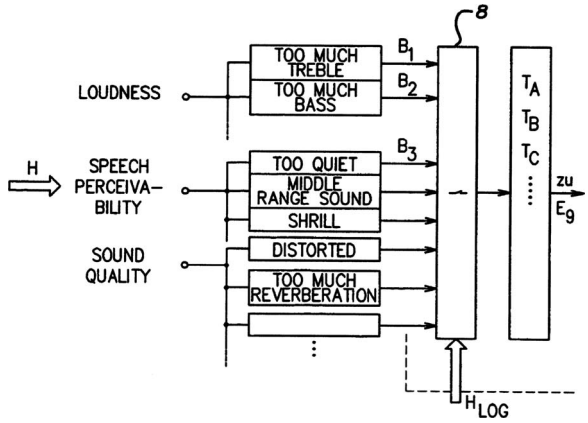
estimated so that a normal vector through the aperture plane aligns with a normal vector in the plane of the second bend of the ear canals. The method applies especially to wireless hearing aids to help ensure good ear-to-ear communication.—DAP

7,450,724

43.66.Ts HEARING AID ADJUSTMENT DEVICE

Michael Greminger, assignor to Phonak AG
11 November 2008 (Class 381/60); filed in Switzerland 17 August 1999

A stored audio test signal, presented from a computerized hearing aid fitting adjustment device, is played from a CD or other audio storage media through a loudspeaker. The wearer's subjective responses to hearing aid fitting parameter variations are gathered via a data entry device and are stored. Subsequent audio test signals are selected automatically, depending on the wearer's responses to the earlier test signals.—DAP

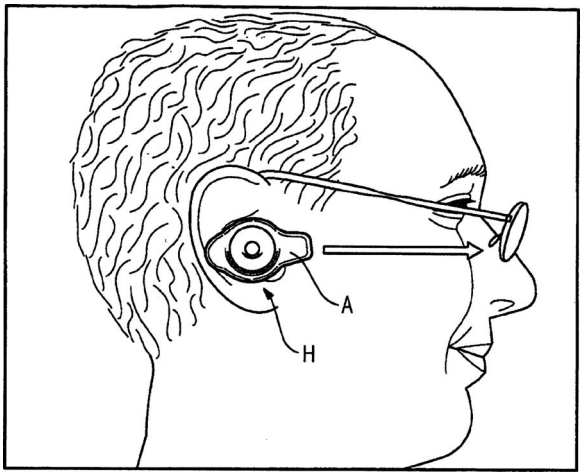


7,450,731

43.66.Ts HEARING AID DEVICE AND CORRESPONDING OPERATING METHOD

Roland Barthel, assignor to Siemens Audiologische Technik GmbH
11 November 2008 (Class 381/313); filed in Germany 4 May 2004

A rotary switch attached to a hearing aid housing selects the operating mode of a directional wireless hearing aid. Hearing aid mode is activated when the directional microphone is pointed with the rotary switching element horizontally. For communication with a telephone, the rotary switch is



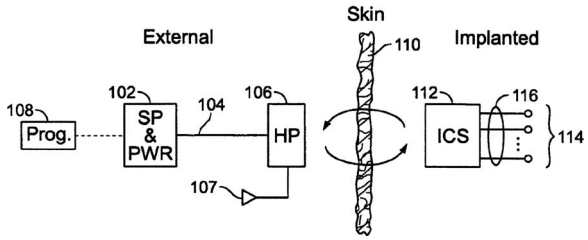
angled downwards toward the wearer's mouth, which automatically powers on and connects a transceiver input and output (e.g., Bluetooth) to the hearing aid. Telephone calls are ended when the rotary switch is returned to horizontal position. The directional microphone and transceiver are housed in a headset-like device.—DAP

7,450,994

43.66.Ts ESTIMATING FLAP THICKNESS FOR COCHLEAR IMPLANTS

Lakshmi Narayan Mishra and Michael A. Faltys, assignors to Advanced Bionics, LLC
11 November 2008 (Class 607/57); filed 16 December 2004

In order to determine an individual optimum magnet strength in the context of how well a cochlear implant microphone/processor module will be held by two magnets in place on the wearer's head behind the ear, the thickness of the skin flap is estimated across which the implanted and external magnets mutually attract. This thickness also is used to determine and adjust the inductive coupling between external transmitter and implanted receiver for through-the-skin communication. Measurements with a skin flap simulator provide reference calibration data.—DAP

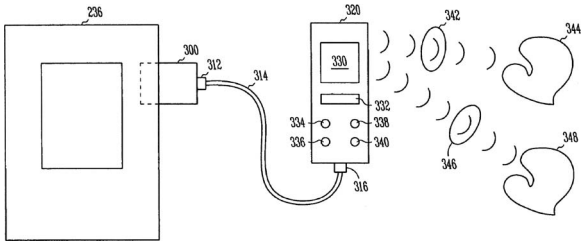


7,451,256

43.66.Ts PORTABLE SYSTEM FOR PROGRAMMING HEARING AIDS

Lawrence T. Hagen and David A. Preves, assignors to Micro Ear Technology, Incorporated
11 November 2008 (Class 710/72); filed 14 January 2005

A non-volatile or volatile memory in a portable device stores hearing aid programming software downloaded from a host computer. A first wireless or wired interface, which is connected to a processor that controls the memory, couples and decouples a hearing aid programmer selectively to a left or right hearing aid so they can be programmed. A second interface, also connected to the processor, couples and decouples the portable device to the host computer to enable downloading the software.—DAP



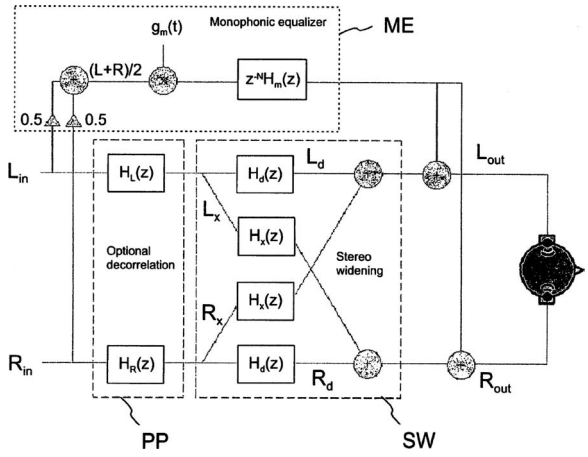
7,440,575

43.72.Gy EQUALIZATION OF THE OUTPUT IN A STEREO WIDENING NETWORK

Ole Kirkeby, assignor to Nokia Corporation
21 October 2008 (Class 381/1); filed in Finland 22 November 2002

The amplitude spectrum of the enhanced monophonic component of the left and right signals of a stereo system intended for headphone listening

is equalized for phase cancellation occurring between the direct path and crosstalk-path signals.—DAP



7,447,640

43.72.Gy ACOUSTIC SIGNAL ENCODING METHOD AND APPARATUS, ACOUSTIC SIGNAL DECODING METHOD AND APPARATUS AND RECORDING MEDIUM

Minoru Tsuji *et al.*, assignors to Sony Corporation
4 November 2008 (Class 704/503); filed in Japan 15 June 2001

The tonal components are extracted from the incoming acoustic time domain signal and are encoded separately for greater accuracy with more bits than are used to encode the remaining components in the signal.—DAP

7,438,685

43.80.Vj APPARATUS AND METHOD FOR REGISTRATION, GUIDANCE AND TARGETING OF EXTERNAL BEAM RADIATION THERAPY

Everette C. Burdette and Dana L. Deardorff, assignors to Computerized Medical Systems, Incorporated
21 October 2008 (Class 600/439); filed 1 November 2002

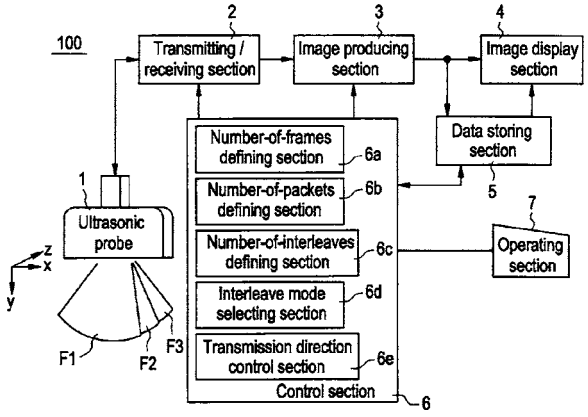
Real-time ultrasound images are fused with radiographic or computed tomography data to localize a region of the prostate for radiation treatment of cancer. Ultrasonic visualization uses a transrectal probe. The approach allows radiation of normal tissue to be reduced and radiation of the abnormal tissue to be increased.—RCW

7,448,997

43.80.Vj ULTRASONIC PULSE TRANSMISSION METHOD AND ULTRASONIC DIAGNOSTIC APPARATUS

Tadashi Shimazaki, assignor to GE Medical Systems Global Technology Company, LLC.
11 November 2008 (Class 600/437); filed in Japan 7 November 2002

Ultrasound pulse transmissions are interleaved in different directions and processed in packets to avoid noticeable blocks of updated information on the display.—RCW



7,448,998

43.80.Vj SYNTHETICALLY FOCUSED ULTRASONIC DIAGNOSTIC IMAGING SYSTEM FOR TISSUE AND FLOW IMAGING

Brent S. Robinson, assignor to Koniklijke Philips Electronics, N.V.
11 November 2008 (Class 600/447); filed 22 February 2005

Ultrasound imaging is performed by alternating between the use of synthetic focusing and conventional beam formation. Images can be formed from harmonic-frequency echoes. Speckle in images can be reduced by combining signals from subapertures that have different beam directions. Regions of interest within images can be processed differently from other images.—RCW

7,450,746

43.80.Vj SYSTEM AND METHOD FOR CARDIAC IMAGING

Fuxing Yang *et al.*, assignors to Verathon Incorporated
11 November 2008 (Class 382/131); filed 26 July 2006

Changes in cardiac chamber volume, wall area, wall thickness, and wall mass are found using three-dimensional ultrasound data acquired at systole and diastole phases of the heart cycle.—RCW

Session 1aAA

Architectural Acoustics: Perception and Binaural Effects in Architectural Acoustics

David H. Griesinger, Chair
23 Bellevue Ave., Cambridge, MA 02140

Contributed Papers

10:00

1aAA1. The importance of the direct to reverberant ratio in the perception of distance, localization, clarity, and envelopment, Part one. David Griesinger (23 Bellevue Ave., Cambridge, MA 02140, dgriesinger@verizon.net)

The component of hall sound that travels directly from a musician to a listener carries information on both the location and distance of the sound source and is vital to both the clarity of the sound and the emotional involvement of listeners. Experiments in laboratories and a hall show that the human ability to separately perceive the direct sound in a reverberant field depends both on the time delay between the direct sound and the reverberant energy and on the direct-to-reverberant ratio (D/R). Threshold data for direct sound perception as a function of time delay and D/R will be presented, and their implications for listening room and hall design will be discussed. In large halls the delay can be long enough that the direct sound is perceptable at D/R values less than −10 dB, but in small halls the threshold becomes as high as 0 dB. Consequences for hall design include: early reflections in the best halls need not be lateral, small halls sound better with fewer early reflections (implying the use of non-shoebox shapes, absorbent stage houses, and shorter reverberation times), and in small halls electroacoustic enhancement should increase only the late reverberation.

10:15

1aAA2. The importance of the direct to reverberant ratio in the perception of distance, localization, clarity, and envelopment, Part two. David Griesinger (23 Bellevue Ave., Cambridge, MA 02140, dgriesinger@verizon.net)

Part two explores the practical consequences of the observations in part one. An experimental method will be presented for objective tests on direct sound detection that yield the threshold data in part one. The rate at which the forward integrated reverberation increases to the level of the direct sound from a short note, and the D/R from short notes or syllables is found to be of particular importance. The insensitivity of current acoustic measures to the perception of direct sound will be discussed, and new measures will be proposed that can be used with live music and speech. We will also present subjective and objective results from a series of experiments in a small hall at the Longy School of Music in Cambridge, Massachusetts. Selective addition of absorption to the stage area and to the lateral walls greatly improved subjective clarity and the balance between musicians on stage and in the hall. Electroacoustics was applied to restore the late reverberation to the value of the unmodified hall without altering the improved clarity. The changes were enthusiastically received by the participants. Audio examples from these experiments will be demonstrated.

10:30

1aAA3. Comparisons of different microphones, microphone orientation, and spacing on late lateral energy measurements. Clothilde Giacomoni, Amy Scherma, and Michelle C. Vigeant (Dept. of Mech. Eng., Acoust. Prog. and Lab., Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117, vigeant@hartford.edu)

A strong correlation exists between listener envelopment and late lateral energy [J. S. Bradley and G. A. Soulodre, "Objective measures of listener envelopment," J. Acoust. Soc. Am. **98**, 2590–2597 (1995)], but limited work has been conducted on measuring this parameter in actual spaces. Measuring late lateral energy (GLL) requires a figure-of-eight pattern microphone to measure the lateral energy only, but an omnidirectional microphone is also used simultaneously to measure other common room acoustic parameters. An investigation was conducted to determine the effects of various different microphones, microphone orientation, and the spacing between the bidirectional and omnidirectional microphone on measured GLL values. Eight figure-of-eight pattern microphones were used, including both ribbon [(1) Cascade Fat Head II, (2) Royer R-121] and condenser [(2) AKG C414, (1) Neumann KM120, (2) Shure KSM44] microphones. Measurements were taken in an 80-seat lecture-style classroom, using the sine sweep method. The microphones were positioned at three different spacings from the omnidirectional microphone, in both orientations at one spacing, and two different receiver locations. In general, differences of 1–2 dB were found between all of the microphones, while minimal differences were found between the two orientations and various spacings for each microphone. Implications of the results for GLL measurements are discussed. [Work supported by University of Hartford WELFund.]

10:45

1aAA4. Effects of room reverberation time and receiver position on measured binaural room impulse responses. Lauren M. Ronsse and Lily M. Wang (Architectural Engr. Prog., Univ. of Nebraska-Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, lronsse@unomaha.edu)

This research examines the effects of room reverberation and receiver position on the fine detail in measured binaural room impulse responses, which can affect psychoacoustic perception in rooms such as speech segregation. Binaural room impulse responses have been gathered in four spaces with a range of shapes and reverberation times for three different receiver positions (center, 1 m from side wall, and 1 m from back wall) and three different source orientations (0, 45, and 90 deg from receiver). The directional loudspeaker is located approximately 0.5 m in front of the receiver for all conditions. Relationships between room reverberation times and measures impacting source localization such as interaural level differences and frequency-to-frequency fluctuations are analyzed. The results will determine if conclusions from previous research in this area conducted in a classroom [B.G. Shinn-Cunningham, N. Kopco, and T.J. Martin, J. Acoust. Soc. Am. **117**, 3100–3115 (2005)] are applicable to spaces with varying shapes and reverberation times.

11:00

1aAA5. Perceptual alignment of early-to-late-reverberant energy ratio and reverberation time to match visual environmental cues of a music performance. Daniel Valente and Jonas Braasch (Commun. Acoust. and Aural Architecture Res. Lab., School of Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180)

The subject of auditory-visual integration has become increasingly important with the introduction of high-resolution playback devices for on-screen object segregation. What effect does the visual location of an object

have on a participant's expectations of that object's auditory stimulus? Several studies have looked at the phenomenon in recent years, but these studies rely on static photographic images to represent the visual scene and mostly use speech signals for testing. Building on these studies, our aim is to propose a testing method using monochromatic compositing (blue-screen technique) to locate a music performance recorded in a studio in a number of virtual acoustical environments being assessed by a listener. In this study, the participant is asked to perceptually adjust two acoustical parameters, early-to-late reverberant energy ratio and reverberation time, of two music performances in five contrasting visual environments according to his or her expectations of how the room should sound given its visual appearance. This study reveals participant resiliency in the presence of forced auditory-visual mismatch; participants are able to adjust the presented acoustic modeling algorithm in a statistically similar way despite randomized starting values for the monitored parameters. Subjective results of the experiment are presented along with objective measurements for verification.

11:15

1aAA6. The impact of visual cues on perceived intimacy in an acoustic ecology. Shane A. Myrbeck, Daniel Valente, Jonas Braasch, and Bobby Gibbs (School of Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, myrbes@rpi.edu)

Crossmodal studies of the effects of visual cues on the perceived acoustics of a performance space have suggested the inextricable relationship between vision and audition. When designing a building or virtual environment for music, critical attention must be paid to a visual aesthetic to ensure a desired level of intimacy between the audience and performer. This visual aesthetic is ordinarily left entirely to the main architect or visual artist. However, given the body of research suggesting crossmodal influences on the human auditory system, it is clear that acousticians must also carefully consider the visual effects of their recommendations for acoustic improvement. The objective of this study is to establish the effects of various visual cues on audition as pertaining to intimacy in an ecological context. This study uses digital compositing (bluescreen) techniques to place a musical performance in various virtual environments, and employs real time audio-video processing to create an interactive, cross-modally immersive environment. Participant-adjusted direct-to-reverberant ratio as well as visual cues adapted from stage lighting techniques are used to establish crossmodal relationships. It is hypothesized that the ability to adjust visual cues will significantly effect subjects' perception of intimate acoustic settings.

11:30

1aAA7. Accurate timbre and localization of binaural recordings through headphones using sound pressure measurement at the eardrum. David Griesinger (23 Bellevue Ave., Cambridge, MA 02140, dgriesiner@verizon.net)

Individual HRTFs are almost universally measured with blocked or partially blocked ear canals, and individual headphone equalization is obtained in the same way. Using this method binaural reproduction without head tracking usually results in a significant alteration of timbre as well as inaccurate and/or in the head localization. The error is important—auralizations using incorrect timbre can lead to misleading conclusions. This problem was studied with the help of a dummy head that precisely models the author's pinna, ear canals, and eardrum resonance. A new type of probe microphone was also developed that allows comfortable recordings and measurements at the eardrum. Data from this equipment show that while the spatial variation in HRTFs can be captured with a blocked ear canal up to a frequency of 7 kHz, errors in headphone equalization made with a blocked canal are typically in error by more than 10 dB at midfrequencies. Eardrum measurements of HRTFs and headphones result in superior performance for an individual, and a pair of headphones correctly equalized for an average individual produces out-of-head localization for at least 50% of other individuals in our tests. Data and theory and a noninvasive method for equalization will be presented.

11:45

1aAA8. Influence of vibration on the perception of musical instruments for various stage constructions. Clemeth L. Abercrombie and Jonas Braasch (Rensselaer Polytechnic Inst., Greene Bldg. 110 8th St., Troy, NY 12180, clem@loudensound.com)

Expanding use of multisensory music presentation calls for increased knowledge of human response to audio-tactile stimulus. This presentation outlines an experiment to explore the human ability to distinguish differences in tactile signals generated by musical sources on typical floor constructions. A double bass and bass drum are used to generate binaural audio and tactile signals on various floor constructions. Vibration signals are then compared for differences in arrival time, magnitude, and frequency content. Audio and vibration signals are reproduced in different combination using a calibrated motion platform and headphones. Test participants are asked if they can identify the differences in tactile signals given a fixed audio environment. Test results are compared to those found in other studies obtained using musical signals as well as harmonic and noise signals. The importance of calibrating tactile musical signals is discussed along with the implications of the results with respect to tactile signal synthesis for multimodal music presentation.

Session 1aABa

Animal Bioacoustics and Psychological and Physiological Acoustics: Fish Bioacoustics: Sensory Biology, Sound Production, and Behavior of Acoustic Communication in Fishes I

Joseph A. Sisneros, Cochair
Dept. of Psychology, Univ. of Washington, Seattle, WA 98195

Richard R. Fay, Cochair
Parmly Hearing Inst., Loyola Univ., Chicago, IL 60626

David Zeddies, Cochair
Marine Acoustics Inc., 4100 Fairfax Dr., Arlington, VA 22203

Chair's Introduction—8:00

Invited Papers

8:05

1aABa1. The future of fish bioacoustics. Arthur N. Popper (Dept. of Biology, Univ. of Maryland, College Park, MD 20742, apopper@umd.edu) and Richard R. Fay (Loyola Univ. of Chicago, Chicago, IL 60626)

In 1993 (Brain, Behav. Evol. **41**, 14–38 (1993)] we evaluated changes in our knowledge of fish bioacoustics since our 1973 review and proposed ten issues that still needed to be answered. This presentation asks whether we have have made progress on these issues since 1993 and poses questions for future fish bioacoustics research. Many of the topics suggested in 1993, such as sound localization mechanism, the roles of the swim bladder' and other peripheral structures in hearing, and the contributions of pressure and particle motion in hearing are still open. There are also "newer" questions not apparent in 1993, such as mechanisms of ultrasound detection and the effects of increased human-generated noise on fish. The usefulness of thinking of fish as "hearing specialists" and "hearing generalists" still needs to be assessed. Finally, an overriding question related to the significance of the diversity of fishes and diversity in ear structure, has not been resolved. Indeed, we still do not know whether the diversity found in the fish auditory periphery and CNS are convergent mechanisms to glean the same kind(s) of information about sound, or whether the variation reflects different uses of sound by different species.

8:20

1aABa2. Fish hearing and bioacoustics since 1973. Richard Fay (Parmly Hearing Inst., Loyola Univ. Chicago, IL 60626) and Arthur Popper (Univ. of Maryland, College Park, MD 20742)

In 1973, we wrote a review in *J. Acoust. Soc. Am.* that summarized our view of fish hearing and bioacoustics. This presentation looks back on the the major questions we had at the time and how the field has dealt with them, or not, to the present. Some of the issues identified included: (1) the roles of the various otolith organs for hearing; (2) lateral line function and its relation to hearing; (3) sound source localization; and (4) how sound stimulation works for fishes without obvious specializations for pressure reception. There have been major advances in these and other aspects of fish sound reception, but in every case, there are significant unanswered questions. For example, it is still controversial whether the utricle plays a role in hearing in most species, and the functional role of the lagena remains an enigma. However, we now have a much better understanding of lateral line stimulation and function, and we now understand that it is not an "accessory auditory organ," as was widely assumed initially. Many questions of sound localization are now thought to be answered theoretically, and in some cases empirically, but many critical and fundamental experiments on localization remain to be done.

8:35

1aABa3. Response of fish sound production to short-term ocean upwelling events at the LEO-15 (Long-term Ecosystem Observatory at 15 m depth) ocean observatory. David A. Mann (Coll. of Marine Sci., USF, 140 7th Ave. S., St. Petersburg, FL 33701, dmann@marine.usf.edu) and Thomas M. Grothues (Rutgers Univ., Tuckerton, NJ 08087)

Understanding factors controlling the distribution and reproduction of fishes is crucial for developing ecosystem models of fish populations. Yet these observations are difficult to make on physically relevant time and space scales. A hydrophone was used to record fish sound production associated with reproduction at the LEO-15 ocean observatory to determine the relationship between variation in fish calling behavior and oceanographic variation. Sound production was dominated by Atlantic croaker (*Micropogonias undulatus*), weakfish (*Cynoscion regalis*), and striped cusk-eel (*Ophidion marginatum*). Striped cusk-eels called with a crepuscular pattern, with a strong peak at dusk, less sound production during the night, and a lesser peak in sound production at dawn. Sciaenids called mostly at dusk and night. Nine advection events bringing colder waters to the LEO-15 site were correlated with greatly reduced levels of sound production in Atlantic croaker and weakfish on daily time scales. Our results show how ocean observatories with passive acoustics can study tightly coupled physical oceanographic influences on fish behavior on daily time scales.

8:50

1aABa4. Stereotypy and variability in the mating call of the Lusitanian toadfish, *Halobatrachus didactylus*. M. Clara P. Amorim, J. Miguel Simoes, Vitor C. Almada (Behavioural Ecology Res. Unit, I.S.P.A., Rua Jardim do Tabaco 34, 1149-041 Lisbon, Portugal), and Paulo J. Fonseca (Univ. of Lisbon, 1749-016 Lisbon, Portugal)

During the breeding season, Lusitanian toadfish males become territorial and emit a tonal advertisement call (the boatwhistle) to attract mates to their nests. We recorded 16 males in the Tagus River estuary (Portugal) during the breeding season and studied boatwhistle variability in a short (minutes) and long time scale (days). We also related boatwhistle acoustic features with male physical characteristics. The recorded males were inside closed artificial nests in groups of eight individuals. Boatwhistles showed individuality in short periods of time but some males showed significant intraindividual variability in a longer time scale. Both social interactions (number of calling males) and tide level changed significantly male calling rate and acoustic characteristics of boatwhistles. The fundamental frequency of boatwhistles (equivalent to the sonic muscle contraction frequency) and boatwhistle emission rate were correlated with male condition (somatic and lipid content). Sound emission rate was also correlated with sonic muscle index. These results suggest that different toadfish males produce distinct boatwhistles, and that they have the ability to modulate the acoustic characteristics of their sounds with the level of social interactions, which increase during rising and high tide. Moreover, boatwhistles characteristics and acoustic activity seems dependent on male condition and sonic muscle index.

Contributed Paper

9:05

1aABa5. Recordings of sonic muscle activity of the Lusitanian toadfish *Halobatrachus didactylus* during sound production. Joana M. Jordão, Maria Clara Amorim, and Paulo Fonseca (Centro de Biologia Ambiental, FCUL, Campo Grande, Lisboa, Portugal)

Male Lusitanian toadfish, *Halobatrachus didactylus*, are territorial and have a wide acoustic repertoire. The sound production apparatus consists of a pair of intrinsic large sonic muscles on the two lobes of the swimbladder. Sonic muscle activity and the variability of the vocalizations were related using electromyograms (EMGs). EMG electrodes were cirurgically implanted on the sonic muscles of several territorial males. After recovery, the subjects were released into an experimental tank with an available nest. Sonic muscle activity was registered and the vocalizations were simultaneously recorded using a hydrophone. Social context was manipulated by releasing different kinds of intruders, male(s) or female(s). This action elicited the production of boatwhistles, while grunts were generated during fish disturbance. Preliminary analysis shows that, unlike previously reported, the bilateral sonic muscles contract simultaneously during the generation of both signal types. The muscle contraction period is halved during the production of the grunt relative to the boatwhistle. The EMG period matches the sound pulse period and determines the fundamental sound frequency. [JMJ was supported by a Post-doc grant (SFRH/BPD/36510/2007).]

Invited Papers

9:20

1aABa6. Toadfish boatwhistles also function as territorial keep-out signals. Raquel O. Vasconcelos (Dept. de Biologia Animal and Centro de Biologia Ambiental, Faculdade de Ciências, Univ. de Lisboa. Bloco C2 Campo Grande, 1749-016 Lisbon, Portugal, rfvasconcelos@fc.ul.pt), J. Miguel Simoes (Unidade de Investigacao em Eco-Etologia, 1149-041 Lisbon, Portugal), Paulo J. Fonseca (Univ. de Lisboa, 1749-016 Lisbon, Portugal), and M. Clara P. Amorim (Unidade de Investigacao em Eco-Etologia, 1149-041 Lisbon, Portugal)

Boatwhistles produced by the toadfish *Halobatrachus didactylus* seem to function as mate attraction calls during the breeding season. Recent observations, however, indicate that their emission is not restricted to this period, suggesting that boatwhistles may have other functions. The possible dual function of boatwhistles was tested by eliciting sound production from reproductive males in advertisement and territorial defence contexts. These were simulated by creating aggregations of confined nesting males in an intertidal area and by conducting territorial intrusion experiments in tanks, respectively. Furthermore, we investigated whether parental care (eggs in the nest) affected the behavioral responses of territorial males. Nesting males kept in aggregations emitted boatwhistles spontaneously for several days. The relation between calling rate and number of eggs in the nest is under analysis. During territorial intrusions, resident males prevented the entrance of intruders in their shelters by producing mostly boatwhistles. Parental males revealed higher aggression levels, exhibiting additional threatening and attack behaviors. Agonistic boatwhistles differed from the mating calls by presenting less amplitude modulation and lower dominant and fundamental frequencies. These results suggest that, apart from attracting mates, the boatwhistles of batrachoidids may also function as an active keep-out signal during territorial defence.

9:35

1aABa7. Evolution and diversity of batrachoidid vocalizations. Aaron N. Rice and Andrew H. Bass (Dept. Neurobiology and Behavior, Cornell Univ., Ithaca, NY 14853, arice@cornell.edu)

Despite the fact that toadfishes are one of the best-studied taxa for understanding fish vocal communication, vocalizations have only been recorded from a low proportion of taxa within the family. Here, we describe the diverse vocal repertoire and novel swimbladder morphology of a tropical Pacific euryhaline toadfish, *Batrachomoeus trispinosus*. *B. trispinosus* produces four classes of vocalizations: long duration hoots, grunts, grunt trains, and grunts exhibiting acoustic beats; the majority exhibiting prominent harmonic features. Calls with acoustic beats, previously unreported for any individual actinopterygian taxon and convergent with songbirds, typically had a predominant frequency around 2 kHz with a beat frequency around 300 Hz. *B. trispinosus* possesses a bilaterally divided swimbladder, forming two completely separate bladders. Grunts exhibiting beats may be produced by bilateral control of the bladders. Phylogenetic analysis based on available molecular data show that *B. trispinosus* and *Halobatrachus didactylus* represent basal lineages in the Batrachoididae, and the separated swimbladders are an independently derived feature of *B. trispinosus*. Thus, the toadfishes may possess higher levels of vocal, motor, and morphological diversity than previously appreciated, and now provide a framework for understanding integrated mechanisms underlying the evolution of acoustic novelty in fishes.

9:50

1aABa8. Sound production and hearing in coral reef butterflyfishes. Timothy C. Tricas and Kelly Boyle (Dept. of Zoology and Hawaii Inst. of Marine Biol., Univ. of Hawaii at Manoa, Honolulu, HI 96822)

Butterflyfishes (family Chaetodontidae) exhibit a diversity of close affiliative social behaviors. All 87 species of *Chaetodon* have a laterophysic connection and extensions of the anterior horns of the swim bladder near the otic capsule, which are proposed to impart sound pressure sensitivity to the normally mechanosensitive lateral line and acceleration-sensitive inner ear. Video kinematic analyses reveal that the motor action patterns used to produce pulsed sounds vary among species. The sound production mechanisms in *Chaetodon* and *Forcipiger* do not appear to be conserved, thus sound production may have evolved independently within the family. Species with long swim bladder horns have lower AEP hearing thresholds than those without, and deflation of the horns and bladder decreases hearing sensitivity in the former group. When the laterophysic connection and swim bladder horns are physically disrupted in wild fish, movement within their territories decreases and fish pairs swim closer together. Thus, the low intensity sounds produced by butterflyfishes may favor close social behaviors that increase communication efficiency among individuals on a noisy coral reef. We are currently investigating how hydrodynamic flow fields produced during social interactions may be used for intraspecific communication during these natural behaviors.

10:05—10:25 Break

10:25

1aABa9. Head and body kinematics of pulse sound generation and feeding in longnose butterflyfishes (Genus *Forcipiger*). Kelly S. Boyle and Timothy C. Tricas (Dept. of Zoology, Hawaii Inst. of Marine Biology, Univ. of Hawaii, 2538 McCarthy Mall, Honolulu, HI 96822, kboyle@hawaii.edu)

Both species of butterflyfish (*Chaetododntidae*) in the genus *Forcipiger* produce pulsatile sounds during territorial interactions. Sounds of one species, the forcepsfish (*F. flavissimus*), are approximately 20 ms and occur with rapid dorsoventral head motion (headbob). We performed laboratory sound production and feeding experiments with high-speed video (600 fps) to test predictions of the hypothesis that sound production in this genus results from evolutionary modification of stereotyped feeding behavior. Cranial elevation is an important component of prey capture for a variety of teleost fishes and facilitates opening of the buccal cavity and oral jaws. Studies on kinematics of feeding in *Forcipiger* by other authors revealed a minor contribution of cranial elevation compared to other fishes. Preliminary results indicate cranial elevation during headbobs involves a comparable range of motion relative to prey capture (approximately 3°–7°), but maximal elevation occurs within 12 ms during the headbob compared to 25 ms in feeding. Additionally, sound emission during feeding is weak to nonexistent. Electroymyography has revealed involvement of anterior epaxialis muscles with cranial elevation during feeding and headbob sound production. These experiments will allow further kinematic comparisons of head movements and associated muscle firing patterns during feeding and acoustic behaviors.

10:40

1aABa10. Sound production and localization in loricariid catfishes. Michael E. Smith, Patrick C. Stewart, Amanda L. Webb, and Brian D. Rogers (Dept. of Biology, Western Kentucky Univ., 1906 College Heights Blvd. #11080, Bowling Green, KY 42101, michael.smith1@wku.edu)

Catfishes of the family Loricariidae have bilobed swim bladders that are adjacent to their ears. We hypothesize that this anatomical design assists these fishes in sound localization and/or sound production. Loricariid catfishes produce short, broadband clicks via pectoral spine stridulation. To test whether these clicks can be localized, we used classical conditioning to train groups of *50 Otocinclus affinis* to come to a conspecific sound on one side of a 200-L aquarium. Four types of experiments were performed—naive, trained, test-1, and test-2. During trained trials, food and a conspecific sound stimulus were presented simultaneously, while only the sound stimulus was presented during naïve (untrained) trials. During the test-1 and test-2 trials, only the conspecific sound stimulus or a 500-Hz sound stimulus, respectively, was presented. *O. affinis* were attracted to the conspecific sound post-training, with 60–80% of the fish remaining on the speaker side of the tank during the stimulus. In summary, *O. affinis* are able to localize a conspecific sound source. Future experiments have been designed to examine the acuity of this sound localization using four speakers in a circular tank, and the effects of swim bladder deflation on the sound localization ability of *O. affinis*.

Contributed Paper

10:55

1aABa11. Bocaccio sounds in the Southern California Bight. Ana Širović (Alaska Pacific Univ., 4101 Univ. Dr., Anchorage, AK 99508, asirovic@alaskapacific.edu) and David A. Demer (Southwest Fisheries Sci. Ctr., NOAA Fisheries, La Jolla, CA 92037)

Sound production by many fish species has been studied extensively, but little is known about sound production by rockfishes (genus *Sebastes*), and only a few species have been reported to be soniferous. Passive acoustic recordings were made during 2007 and 2008 at Hubbs-SeaWorld Research Institute and Southwest Fisheries Science Center tanks containing Bocaccio (*S. paucispinis*). Sounds in tanks were recorded using preamplified hydrophones (HTI-94-SSQ) and were digitized at sample rates of 44,100 or 8000 Hz. Recordings were collected at 14 locations in the Southern California Bight (SCB) using passive acoustic moorings, AURALs, sampling at 8000 Hz from August through October 2007. Three low frequency (<900 Hz), short duration (<4 s) sounds were recorded in tanks containing only Bocaccio multiple times. One of these sounds was also commonly recorded in the SCB and is the same sound that was recorded first by Thompson off San Clemente Island in the 1960s. This sound exhibited the same diel pattern as noted by Thompson, with maximum calling at night, peaking at dusk and dawn. This long-term persistence of the same sound indicates that passive acoustic tools may be used for efficient monitoring of changes in rockfish populations over long time periods.

Invited Paper

11:10

1aABa12. Sound source localization by the plainfin midshipman fish, *Porichthys notatus*. David Zeddies (Marine Acoust. Inc., 4100 Fairfax Dr., Arlington, VA 22203), Richard R. Fay (Loyola Univ. Chicago, Chicago, IL 60626), Peter W. Alderks (Univ. of Washington, Seattle, WA 98195), Kiel Shaub (Univ. of Washington, Seattle, WA 98195), and Joseph A. Sisneros (Univ. of Washington, Seattle, WA 98195)

Sound source localization of the plainfin midshipman fish (*Porichthys notatus*) was studied using the phonotactic response of gravid females to synthetic male advertisement calls. Playback experiments were conducted in a 3.65-m-diameter outdoor concrete tank at the UC Bodega Marine Laboratory in Bodega Bay, CA using a J-9 transducer placed near the center of the tank. The sound field in the tank was measured at 5-cm intervals using an eight-hydrophone array to measure the pressure gradients from which particle motion vectors were calculated. The acoustic measurements confirmed that the J-9 projector was operating as a monopole source. One hundred twenty-two gravid females were released 100 cm away from the center of the sound source while their behavioral responses were video taped during experiments with playback sound "On" (90 Hz tone) and sound "Off" (Controls). A total of 45 positive phonotactic responses from naïve gravid females were recorded from the 62 trials with sound "On." The phonotactic responses consisted primarily of straight to slightly curved tracks to the sound source while abrupt changes in trajectory to the sound source were rarely observed. The results confirm that fish can locate sound sources in the near field. [Work supported by NSF.]

Contributed Paper

11:25

1aABa13. Passive acoustic determination of herring size. Thomas R. Hahn, Jennifer Wylie, and Xiaojun Chen (Rosenstiel School of Marine and Atmospheric Sci. Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

Over the last decade, acoustic signatures of a variety of fishes have been recorded and analyzed. More recently, these vocalizations have been used to passively detect aggregations of fish, demonstrating the potential to supplement and enhance traditional active acoustic surveys. Based on previously published work on acoustic emissions of herring, this paper discusses the possibility of not merely passively detecting absence or presence of aggregations of herring but, additionally, assessing abundance and size. Theoretical considerations as well as data collected in Prince William Sound, Alaska, on Pacific herring (*Clupea Pallasii*) are presented.

Invited Paper

11:40

1aABa14. Assessment of cumulative sound exposure level as criterion for exposure of fish to impulsive sound. Michele B. Halvorsen (Dept. of Biology, Univ. of Maryland, College Park, MD 20742, mb_halvorsen@yahoo.com), Thomas Carlson (Pacific Northwest Natl. Labs., Battelle, Portland, OR 97204), and Arthur N. Popper (Univ. of Maryland, College Park, MD 20742)

Acceptable exposure of fish to pile driving impulsive sound is currently determined by cumulative sound exposure level (SEL). The implicit assumption in this regulatory approach is that cumulative SEL, regardless of how reached, is one of the critical measures of sound exposure for fish health. We are testing this assumption by exposing juvenile Chinook salmon to sequences of impulsive sound differing in individual impulse SEL and number of impulses, but with the same cumulative SEL value. Test fish are exposed in a unique device that permits accurate simulative of impulsive pile driving sound. Following exposure, the hearing of test fish is evaluated using auditory evoked potential methods followed by necropsy for assessment of barotrauma. The null hypotheses we are testing is that the probability of hearing threshold shift $\geqslant$6 dB and the probability of barotrauma mortal injury is the same for all cumulative SEL exposures. The results of these tests will determine the experimental design of additional tests to derive criteria for exposure of fish to pile driving sound. [Work supported by National Cooperative Highway Research Program (NCHRP) and Minerals Management Service (MMS).]

Session 1aABb

Animal Bioacoustics: An Integration of Bioacoustics, Neuronal Responses, and Behavior I

Terry Takahashi, Chair
Inst. of Neuroscience, Univ. of Oregon, Eugene, OR 97403

Chair's Introduction—8:00

Invited Papers

8:05

1aABb1. Spatial aspects of acoustic processing in complex environments. H. Steven Colburn, Barbara Shinn-Cunningham, and Kamal Sen (Hearing Res. Ctr., Boston Univ., 44 Cummington St., Boston, MA 02215, colburn@bu.edu)

From the perspective of acoustical signal processing, it is a challenge to analyze a complex sound environment and segregate the acoustic mixture into individual sources, each with its own spatial characteristics and reverberation pattern. However, auditory systems perform this perceptually critical task with impressive ease. This talk will review how acoustical array processors work to separate sound sources. We then consider what most vertebrates, who have two point receivers separated in space, can accomplish by combining the signals that reach their two ears. The nature of the information available to bird and mammalian auditory systems is briefly reviewed, and the computational approaches these systems take to separating sound sources is assessed, including the logic of processing sound within narrow frequency channels over brief time epochs (rather than, e.g., undertaking broadband analysis). Some observations about the ability of classical binaural processing models to implement earlier stages of processing for these complex environments will be discussed. Finally, performance of human and zebra-finch subjects in tasks requiring judgments about individual sources embedded in complex sound mixtures will be considered in the context of physiologically motivated processing schemes. [Work supported by NIDCD: DC00100 (Colburn), DC05778 (Shinn-Cunningham), and DC07610 (Sen).]

8:30

1aABb2. Neural basis of spatial hearing in reverberant environments. Sasha Devore (Eaton Peabody Lab., Boston, MA 02114 and Harvard-MIT Div. of Health Sci. and Technol., Cambridge, MA 02139), Antje Ihlefeld, and Barbara Shinn-Cunningham (Boston Univ., Boston, MA 02215), and Bertrand Delgutte (Eaton Peabody Lab., Boston, MA 02114)

In reverberant environments, acoustic reflections interfere with the direct wavefront reaching a listener's ears, distorting the spatial cues for sound localization. Yet, human listeners have little difficulty localizing sounds in everyday settings. Our research aims to elucidate the neural basis of spatial hearing in reverberant environments. Using a virtual acoustic space (VAS) technique, we investigated the effects of reverberation on the directional sensitivity of binaural neurons in the inferior colliculus (IC) of anesthetized cats and awake rabbits, focusing on neurons sensitive to interaural time differences. Consistent with the buildup of reverberant energy in a sound stimulus, we find that reverberation increasingly degrades the directional sensitivity of single neurons over time, although the amount of degradation depends on the characteristic frequency and the type of binaural cues available. The neurophysiological data can account for results from recent human psychophysical studies of spatial hearing in reverberant environments. In particular, we show how a population rate model for decoding the observed midbrain responses predicts the main trends in human lateralization performance, suggesting a subcortical origin for robust sound localization in reverberant environments. [Work supported by National Institutes of Health (NIH) Grant Nos. R01 DC002258 (BD), P30 DC005209 (BD), and R01 DC05778-02 (BGSC). SD partially supported by NIH Grant No. T32 DC00038 and the Helen Carr Peake Fund.]

8:55

1aABb3. Comparative studies of the cocktail party problem using birdsong. Micheal L. Dent (B76 Park Hall, Dept. of Psych., Univ. at Buffalo, SUNY, Buffalo, NY 14260), Barbara G. Shinn-Cunningham, and Kamal Sen (Boston Univ., Boston, MA 02215)

Many animals are adept at identifying communication calls in the presence of different masking sounds, much like humans navigating a cocktail party. In order to compare human and avian abilities in analyzing signals in complex acoustic mixtures, we measured identification of birdsong by humans and birds in the presence of different maskers and compared results to physiological recordings in zebra finches. In humans, spectrotemporal similarity of targets and maskers causes difficulties in segregating and selecting the target out of the sound mixture; however, for similar targets and maskers, spatial separation alleviates this interference. Field L recordings in finches reveal differences in the forms of neural interference caused by the various masker types, consistent with human behavior. However, avian identification performance is similar for the different maskers. Moreover, spatial separation of target and masker causes little improvement in avian target identification unless target and masker have distinct temporal envelopes. When target and masker envelopes differ, avian identification performance improves significantly, regardless of masker type. These experiments demonstrate that both birds and humans can segregate communication signals from a sound mixture, but suggest that the acoustic attributes that the two species use to solve the cocktail party problem differ.

9:20

1aABb4. Localization and identification of concurrent sounds in the owl's auditory space map. Clifford H. Keller and Terry T. Takahashi (Inst. of Neurosci., Univ. of Oregon, Eugene, OR 97403)

In nature, sounds from multiple sources sum at the eardrums, generating complex cues for sound localization and identification. In this clutter, the auditory system must determine "what is where." We examined this process in the auditory space map of the barn owl's inferior colliculus using two spatially separated sources simultaneously emitting uncorrelated noises, which were uniquely identified by different frequencies of sinusoidal amplitude modulation. The neurons responded when either source was placed within the receptive field, suggesting two clearly segregated foci of activity at appropriate loci on the space map. The spike trains were locked strongly to the AM of the source within the receptive field, while the other source had minimal influence. Two sources AMed at the same rate were successfully resolved, suggesting that source separation is based on differences of fine structure. The spike rate and synchrony were stronger for whichever source had the stronger average binaural level. A computational model showed that neuronal activity was largely proportional to the degree of matching between the momentary binaural cues and the neuron's preferred values. Individual neurons respond to and synchronize with sources in their receptive field if there are frequencies having an average binaural level-advantage over a second source. [NIH-DC003925.]

9:45

1aABb5. Localization of amplitude modulated sounds and their echoes in the barn owl. Brian S. Nelson and Terry T. Takahashi (Inst. of Neurosci., Univ. of Oregon, Eugene, OR 97403)

Echoes that arrive after short delays are not perceived as events separate from sounds arriving directly from an active source, a phenomenon known as localization dominance. Studies have focused on the time-advantage at the onset of an echo/direct-sound pair to explain localization dominance. The head-saccades of barn owls and the responses of neurons in their auditory space maps were measured when sound pairs were gated simultaneously, so that the lead/lag relationship could be discerned only from the ongoing features of the envelopes. The owls behaved as though they experienced localization dominance at short delays, but only if the sounds were deeply amplitude modulated, suggesting the importance of the ongoing envelope. Correspondingly, neurons in the space map fired preferentially to the leading sound, thus paralleling the behavior, but again, deep amplitude modulations were necessary. Analyses of waveforms at the ears suggest that the acoustic features that caused the space map neurons to fire were more abundant for the leading than for the lagging stimuli, thus explaining localization dominance without the need for explicit echo-suppression circuitry. [Work supported by NIDCD F32-DC008267 and RO1-DC003925.]

10:10—10:30 Break

10:30

1aABb6. Coding of sound features in primary and secondary auditory areas of songbirds. Frederic E. Theunissen (Dept. of Psych., Univ. of California Berkeley, 3210 Tolman Hall, Berkeley, CA 94720)

We have used a combination of neuroethological and classical auditory neurophysiological approaches to study how behaviorally relevant sounds are processed in the avian auditory system. One of the contributions of the neuroethological approach has been the discovery of highly specific auditory neurons that appear to be specialized for detecting very specific behaviorally relevant sounds. On the other hand, many auditory neurons recorded in the auditory system of nonspecialists do not exhibit such specificity. At the same time, animals and humans hear and process a large space of sounds and are able to categorize these in much broader perceptual terms, describing them in terms of their pitch, timbre, and rhythm. By systematic analyzing neural responses to song in the ascending avian auditory system and relating receptive fields to the statistics of natural sounds, we have shown that these two approaches can be unified: we found that the spectrotemporal receptive fields of auditory neurons tile a subset of the acoustical space that is particular important for natural sounds. In addition, we found that neurons could be classified into functional clusters. Neurons in different clusters were sensitive to different song features and, we will argue, are involved in mediating distinct perceptual attributes.

10:55

1aABb7. Sound localization acuity in the cat: Integrating data from psychophysical, physiological, and acoustical cue studies. Daniel J Tollin (Dept. of Physio., Univ. of Colorado Med. Sch., 12800 E. 19th Ave., Aurora, CO 80045, daniel.tollin@uchsc.edu)

The minimum audible angle (MAA), a psychophysical measure of auditory spatial acuity, is the smallest angle separating the locations of two sound sources, presented sequentially, that can be reliably discriminated. In humans and cats, the MAA for tone stimuli depends jointly on source azimuth and tone frequency. It is hypothesized that this results from the spatial and frequency dependence of the acoustical cues to location. Thus, for high-frequency ($>$1.5 kHz) stimuli, where the predominant cue for source azimuth is the interaural level difference (ILD), the MAA would be determined by the spatial and frequency dependence of the ILD cue. However, for this to be true, the underlying neural acuity for ILD cues must be frequency independent. Here the hypothesis was tested that the frequency dependence of the psychophysical MAA in cats can be explained by the frequency dependence of the ILD cue with source azimuth and the frequency independence of the acuity of neurons (or neural MAA) in the lateral superior olive (LSO) to ILDs. The LSO is a brainstem structure responsible for the initial encoding ILDs. The data from acoustical cue measurements and physiological and psychophysical studies in cats support this hypothesis. [Work supported by NIH Grants Nos. DC006865 and DC02840.]

11:20

1aABb8. Some brain mechanisms for auditory scene analysis. John C. Middlebrooks, Chen-Chung Lee (Dept. of Otolaryngol., Univ. of California, Irvine, CA 92697-5310, j.midd@uci.edu), and Ewan A. Macpherson (Univ. of Western Ontario, London, ON Canada, N6G 1H1)

Detection and discrimination of sounds in complex auditory environments is facilitated by knowledge of the locations of sound sources and by spatial separation of signals and maskers. In studies of the spatial sensitivity of auditory cortical neurons in anesthetized cats, we find that location-dependent variation in temporal spike patterns can signal sound-source locations throughout up to 360 deg of space. In awake cats, spatial receptive fields can be similarly broad, although cortical neurons show a greater diversity of temporal patterns than is seen in anesthetized conditions. Our recent results demonstrate that spatial tuning tends to sharpen when a cat is engaged in an auditory task, compared to idle conditions, and that tuning can sharpen even further when the task involves sound localization. In human psychophysical studies, spatial hearing permits perceptual "stream segregation" of multiple temporally interleaved sequences of sounds. In parallel studies of the auditory cortex in anesthetized cats, we find that interleaved sound sequences, in configurations that likely would be segregated perceptually, result in activation of discrete cortical neuron populations. The spatial acuity for this neuronal stream segregation is substantially sharper than has been observed in studies of cortical representation of single sound sources.

MONDAY MORNING, 18 MAY 2009 BROADWAY I/II, 10:25 A.M. TO 12:00 NOON

Session 1aAO

Acoustical Oceanography, Underwater Acoustics, and Signal Processing in Acoustics: Temporal and Spatial Field Coherence Applied to Ocean Sensing: Measurement, Theory, and Modeling I

Timothy F. Duda, Chair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543-1053

Chair's Introduction—10:25

Invited Papers

10:30

1aAO1. On the use of broadband acoustic scattering techniques for observing physical and biological scattering features in the ocean. Andone Lavery (Woods Hole Oceanograph. Inst., Dept. of Appl. Ocean Phys. and Eng., MS11, Woods Hole, MA 02543)

Broadband acoustic scattering techniques, and associated signal processing techniques, are emerging as a powerful tool for the remote sensing of physical and biological scattering features in the ocean. These techniques allow (1) the frequency spectrum of oceanic scatterers to be measured, thereby resulting in enhanced classification and quantification potential relative to more traditional single-frequency scattering techniques, and (2) increased spatial resolution resulting in high-resolution images of oceanic scattering features, obtained through use of pulse compression signal processing techniques that exploit the broadband content of the signal. In this presentation, broadband processing techniques are discussed and results from laboratory and field measurements of broadband acoustic scattering from both physical and biological scatterers, spanning the frequency range from 100–600 kHz, are presented. Controlled laboratory broadband scattering measurements allow coherent ping-to-ping analysis to be performed as well as individual scattering features to be determined. The laboratory approaches are then expanded into the field to investigate physical and biological scattering features in the ocean on relevant temporal and spatial scales. Examples of broadband field measurements are presented from recent field experiments, and include scattering from nonlinear internal waves, zooplankton, turbulent oceanic microstructure, salinity microstructure in rivers, and bubbles.

10:50

1aAO2. Coherence in long-range deep ocean propagation. Michael G. Brown, Francisco J. Beron-Vera (RSMAS, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149), Ilya A. Udovydchenkov, and Irina I. Rypina (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Acoustic wavefield structure, stability, and coherence properties in deep ocean environments will be discussed. Properties of transient wavefields will be emphasized. Observations, simulations, and some theoretical considerations will be presented. Some important experimental milestones, beginning approximately in 1979, will be noted. It will be argued that the last 30 years have been characterized by a number of pleasant surprises relating to the stability and coherence of transient wavefields in deep ocean environments. [Work supported by ONR.]

11:10

1aAO3. Transverse coherence lengths, processing limits and implications. William M. Carey (Dept. of Mech. Eng., College of Eng., Boston Univ., Boston, MA 02215)

How does scattering in sound channels (deep and shallow waters) limit coherent array processing or what is the limitation of resolution in terms of the mutual coherence function and its temporal and spatial coherence lengths? The resolution of an array is limited by the mutual coherence function; but estimation in a partially coherent noise background with a multipath signal is difficult using the normalized cross power spectral density, magnitude squared coherence, because of the properties of both signals and noise. The measurement of magnitude-squared coherence is a poor statistical estimator since it is a function of the signal-to-noise ratio and multipath interference with large confidence bounds. Array gain measurements and a wave-theoretic coherence functional form can provide estimates of temporal and spatial coherence lengths defined as the 1/*e* value of this function. This paper reviews single path coherence results and those derived from array measurements over the low- to mid-frequency range in deep and shallow water. Representative coherence lengths are discussed in terms of boundary interactions, internal wave scattering, and coastal mesoscale features. The implications for arrays used to estimate geoacoustic properties, mammal locations, and scattering from the boundaries are presented.

Contributed Papers

11:30

1aAO4. Comparing the effects of internal waves, mode coupling, and change in bandwidth of a radiated signal on low mode energy propagation. Natalie Grigorieva, Gregory Fridman (Dept. of Appl. Math., St. Petersburg State Marine Tech. Univ., 3 Lotsmanskaya Str., St. Petersburg 190008, Russia, nsgrig@natalie.spb.su), James Mercer, Rex Andrew, Michael Wolfson (Univ. of Washington, Seattle, WA 98105), Bruce Howe (Univ. of Hawaii, Honolulu, HI 96822), and John Colosi (Naval Postgrad. School, Monterey, CA 93943)

Wideband signal propagation modeling is carried out for the actual space-time configuration realized during the long-range ocean acoustic propagation experiment (LOAPEX) conducted in the North Pacific in 2004. In particular, the experiment used the Kauai transmit station that was located at a range of 2432 km from a vertical line array (VLA), M-sequence at 75 Hz, and a source depth of 800 m close to the depth of a sound-channel axis. Two sound speed profiles were utilized to get the smooth two-dimensional sound speed field. The first one was the profile obtained from the conductivity-temperature-depth (CTD) measurements at the Kauai station. The second sound speed profile from the VLA location was based on a CTD cast taken at station T50 (50 km from the VLA). To take into account the acoustic fluctuations due to internal waves, the buoyancy frequency profile based on LOAPEX CTD measurements made at station T50 was used. For simulations, 512 values of the horizontal wave number were utilized with the maximal mode number equal to 50. The low modes intensity received by the hydrophones close to the sound channel axis at the VLA is compared with the results of simulations. [Work supported by ONR and ONR Global.]

11:45

1aAO5. On the time-mean state of ocean models and the properties of long-range acoustics. Brian D. Dushaw (Appl. Phys. Lab., Univ. of Washington, 1013 N.E. 40th St., Seattle, WA 98105, dushaw@apl.washington.edu), Matthew A. Dzieciuch, and Peter F. Worcester (Univ. of California at San Diego, La Jolla, CA 92093-0225)

Global and regional circulation models of the ocean have the horizontal and vertical resolution required for realistic simulation of long-range acoustic propagation. These simulations offer a novel test of model accuracy, because acoustic propagation is sensitive to the vertical gradients of sound speed in the models and acoustic arrival patterns are known to have stable and universal properties. The travel time and dispersal of an arrival pattern are examples of such properties. Acoustic receptions on three long vertical line arrays from basin-scale transmissions in the North Pacific in 1996 and 1998 are used to test the acoustical properties of the time-mean state of several ocean models. The NOAA World Ocean Atlas, a global representation of the "average" ocean, yields acoustic predictions whose patterns closely match the observations, but not all ocean models accurately represent oceanic sound speed properties. Acoustical tests of a model do not necessarily require actual data, because basic acoustical properties may be derived from the World Ocean Atlas.

MONDAY MORNING, 18 MAY 2009 PARLOR B/C, 8:00 A.M. TO 12:00 NOON

Session 1aNS

Noise and Architectural Acoustics: Prediction and Control of Noise Related to Buildings

James E. Phillips, Chair
Wilson Ihrig & Assoc Inc., 5776 Broadway, Oakland, CA 94618-1531

Contributed Papers

8:00

1aNS1. Research into the human response to vibration in residential environments. David C. Waddington, Andrew T. Moorhouse (Acoust. Res. Ctr., Univ. of Salford, Salford M5 4WT, UK, d.c.waddington@salford.ac.uk), and Phil A. Brown (Univ. of Salford, Salford M5 4WT, UK)

This paper describes work being carried out at the Univ. of Salford to develop a method by which human annoyance to vibration in residential environments can be assessed. The objective of this study is to yield a robust relationship between vibration exposure and human response; therefore, providing a reliable basis for the development of standards and guidance for the assessment of vibration in residential buildings. The vibration sources to be considered are those affecting residents that are outside their control, such as construction, road, and rail activities. The protocol involves the measurement of vibration outside and inside residences and a social study questionnaire based on face-to-face interviews with adults. The project is expected to capture a wide-scale study of 2000 case studies, and examples of early field measurements and results are presented. This work is likely to be of interest to planners and practitioners involved in the design of buildings and environmental health officers involved in the assessment of vibration complaints. [Work funded by the Dept. for Environment, Food, and Rural Affairs (Defra) UK.]

8:15

1aNS2. Comparison of tapping machines for impact insulation class (IIC) measurement accuracy. Aaron M. Farbo (Cavanaugh Tocci Assoc., 327F Boston Post Rd., Sudbury, MA 01776, afarbo@cavtocci.com)

This study compares tapping machines from different manufacturers and age for impact insulation class (IIC) measurement accuracy. In addition to the IIC measurement accuracy, the study also compares the impact frequency and effect of different impact surfaces (steel and rubber). The tapping machines were tested in a model condo unit in accordance with impact insulation class testing standards ASTM E1007-04 and ISO 140-6. These standards clearly state the impact frequency and type of impact surface to be used (hardened steel). Previous testing standards did not specify the material type for impact surfaces and many older tapping machines included both steel and rubber feet. The results of this study will be presented to show the effect of age, impact frequency, and impact surface on the resulting IIC measurement for the tapping machines.

8:30

1aNS3. Investigation of damping materials for heavy-weight floor impact sounds in concrete structures. Jin Yong Jeon and Seung Yup Yoo (Dept. of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea)

Multi-layers of surface materials were investigated for the isolation of floor impact sounds in concrete structures. Both resilient and viscoelastic damping materials were tested according to the modified method of ANSI S2.22 for calculating elastic modulus and loss factor. The dynamic characteristics of the concrete structures subjected to heavy-weight floor impacts were modeled by the compressional damping theory, using sandwiched beams. Results show that the viscoelastic damping materials are more effective than the resilient materials. In field measurements, the insertion loss of the heavy-weight impact sounds, which was mainly affected by the modal characteristics at low frequencies, was higher in floor structures with the damping materials.

8:45

1aNS4. Ceiling details in wood frame construction. James Keene (Keene Bldg. Products, 5910 Landerbrook Dr., Mayfield Hts., OH 44124, jrk @keenebuilding.com)

In multifamily projects, the details of floor/ceiling installation have been a known and constant problem. Errors when installing noise control elements have caused lots of failures. Specifically, gypsum board ceiling and wall isolation with caulking is often short circuited when the gaps are incorrect. This paper details a new method of installing gypsum board ceilings and walls to provide a more easily inspected and detail assured isolation system. The system applies a similar system for ceiling installation as has become the standard in floor perimeter isolation. An "L" shaped fabric strip is installed around the perimeter of the room to create the gaps that today are created by the contractor in the field. The paper and presentation will review the methodolgy of the system and the laboratory test results that show performace compared to the present approach. The fire rated test performance will also be summarized. Since this is multifamily, the economic savings of this system are also presented.

9:00

1aNS5. Comparison of apparent field sound transmission class (AFSTC) and outdoor indoor transmission class (OITC) as applied to modular classroom design and field analysis. 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, nphilipp@mail.unomaha.edu)

Regarding the drafted addendum for ANSI S12.60-2002 Standard on Classroom Acoustics Guideline, it was recently proposed to utilize Outdoor Indoor Transmission Class as a means of rating modular classroom exterior envelope noise reduction performance. This paper provides a comparative analysis of Outdoor Indoor Transmission Class with the Apparent Field Sound Transmission Class. Measurements were conducted in modular classrooms currently in use by the Omaha Public School District in Omaha, NE, along with newer construction units prior to installation on-site. Limitations to both measurement methodologies are discussed in relation to their source noise spectrums.

9:15

1aNS6. Acoustical analysis of an indoor test facility for a 30-mm Gatling gun. Matthew D. Shaw and Kent L. Gee (Dept. of Phys., Brigham Young Univ., N283 ESC, Provo, UT 84602, mdshaw16@gmail.com)

The Air Force commissioned a construction and engineering company to build an indoor test facility for the GAU-8 Avenger at Hill Air Force Base in Layton, UT. The blast pressures from this 30-mm Gatling gun, however, are large enough to cause spalling of the concrete walls over time. The facility is being designed and constructed to last for 20 yrs, requiring several acoustical treatments. The pressures from the gun were measured outdoors, with maximum pressures exceeding 3000 Pa (163 dB) at a distance of 30 ft (9.1 m). A computer model of the room was designed using EASE, and impulse responses were generated at several positions. These impulse responses were convolved with an ideal blast wave pulse train to mimic the sound of the gun in the room. From these data and results collected from preliminary tests in the range, recommendations have been provided as to placement and types of necessary treatments. [Work supported by HHI Corporation.]

9:30

1aNS7. The effective bandwidth of the enhanced transmission loss across a wall with Helmholtz resonators. Joel Garrelick (Appl. Physical Sci., Inc., 49 Waltham St., Lexington, MA 02421, jgarrelick@aphysci.com)

It is long known that the introduction of Helmholtz resonators can enhance the transmission loss of a wall in the vicinity of their natural frequencies. Resonance is typically achieved by creating open-mouthed cavities in the wall. The effective bandwidth of the phenomenon is a key factor in its practical utility. It is dependent on the internal resonator resistance as well as the *in situ* external acoustic impedance at the cavity mouth. This paper is an analysis of the bandwidth for a variety of resonator configurations, including cavities with septa and differing volumes. The mathematical model, driven by a normally incident plane wave, will treat the wall as rigid but moveable, and the cavities will be infinitely periodic arrays. Influences of wall impedance, resonator resistance, cavity array spacing and natural frequencies will be described. In addition, the potential for minimizing the deleterious effects of diffraction between neighboring cavities by placing the cavity mouths on both sides of the wall will be assessed.

9:45

1aNS8. Vertical profile of community noise in high-rise built environment. Alam Sheikh Mahbub, Lee Siew Eang (Dept. of Bldg., Natl. Univ. of Singapore, Singapore), and W. L. H. Johnny (Housing and Development Board)

High-rise public housing apartments in Singapore are often subjected to different types of community noise sources in close proximity. Community noise in housing estates comprises noise sources such as food center noise, children playground noise, soccer court noise, basketball court noise, waste disposal truck noise, etc. It was found as one of the major noise source categories in residential environment. With the enhanced performance quality in many aspects of public housing in recent years, it is imperative that the acoustics performance of the apartments can be enhanced to match overall quality of public housing environment through appropriate characterization of these noise sources. A scientific and reliable approach for characterizing noise propagation from these sources is vital in the implementation of projects. Software modeling and simulation were carried out to study noise propagation characteristics of five types of community noise sources. Five 16-storey residential buildings were investigated with noise measurement from these sources, conducted on each level of the buildings. The predicted results were verified with field measured data. This study observes that noise level from these sources reduces between 3.7 and 12.8 dBA at a height between 43 and 46.5 m above the ground with the increase in building height.

10:00

1aNS9. Prediction of the reverberation times and sound pressure levels in town streets. Kai Ming Li (Ray W. Herrick Lab., Purdue Univ., 140 S. Martin Jischke Dr., W. Lafayette, IN 47907-2031) and Chenly Y. C. Lai (Hong Kong Polytech. Univ., Kowloon, Hong Kong)

The current study was devoted to investigate the propagation of noise along a town street where building façades and road surfaces form a street canyon. An energy based approach was used to estimate reverberation times and the sound pressure levels due to a point source located along the town street. An image source method was developed to compute sound energy in this semienclosed space. However, the image source model can be simplified leading to the numerical solutions expressed in an integral formulation. The absorption coefficients of the boundary surfaces and the air absorption factor can be included in the present analysis. The integral can further be evaluated analytically with the sound pressure levels expressed in a closed form solution. Both outdoor full scale measurements and indoor scale model experiments were conducted and compared with the computed results. The numerical results due to the proposed formula not only agree well with the experimental measurements but were also in accord with the published data.

10:15—10:30 Break

10:30

1aNS10. Modeling sound transmission through large rectangular openings in a damped enclosure. Buye Xu and Scott D. Sommerfeldt (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

The understanding of sound transmission through large rectangular openings in a damped enclosure plays an important role in designing noise reduction solutions where enclosures and openings are often involved. A complete solution for this problem should consider diffraction, coupling between exterior and interior sound fields, and damped boundaries, each of which is a challenging task. A theoretical model will be introduced in this paper that deals with all these phenomena and, in addition, also considers the effects of the direct sound field in the solution. Numerical simulations and experimental results will be presented and analyzed.

10:45

1aNS11. Prediction of sound propagation in rectangular ducts by a ray-tracing model with phase. Behrooz Yousefzadeh and Murray Hodgson (Acoust. and Noise Res. Group, Univ. of British Columbia, 3rd floor, 2206 East Mall, Vancouver, BC V6T 1Z3, Canada, behroozy@interchange.ubc.ca)

A ray-tracing model which includes phase changes due to distance traveled and surface reflection has previously been developed for predicting sound field in rectangular rooms. Including phase has removed the limitation associated with energy-based models, that room dimensions must be larger than the sound wavelength. The model has been applied to a rectangular room with dimensions not large compared to the wavelength. The room dimensions are that of a real-world ventilation duct, with the boundary conditions idealized as a point source at one end and anechoic termination at the other end. Predictions have been made of the pressure field inside the duct and the results have been compared with analytical solutions. The model is found to be capable of predicting the modal characteristics of the duct.

11:00

1aNS12. Assessing the origin of tonal sound emanating from refrigerant piping in the historic Mattress Factory project at the University of Washington Tacoma Campus. Elizabeth Bogomolov and Julie A. Wiebusch (The Greenbusch Group, Inc., 1900 West Nickerson St., Ste. 201, Seattle, Washington 98119)

UW Tacoma campus is located in the Union Depot Warehouse National Historic District. Siting the campus in this area and renovating rather than demolishing some of the remaining historic, industrial buildings as campus classrooms and offices has stimulated the city's urban renewal. Mattress Factory was one of these historic structures, regulated by the Tacoma Landmarks Commission. The new chilled water system in Mattress Factory consists of two 70-ton chillers located in a mechanical room on the lowest level of the four-story building. Air-cooled condensing units, located on the roof, connect to the chiller compressors. The refrigerant piping was required to be incorporated within the building envelope so as not to alter the appearance of the historic exterior. The piping path between the compressors and condensers was somewhat circuitous due to existing structural elements. A discernable, tonal whine was exhibited in spaces on the second and third floors when the chiller was running. The sound corresponded with the location of the refrigerant piping within the wall cavity, with the dominant sound level at the second floor in the area of an existing beam. This paper outlines the process of identifying the cause of the sound and the ultimate mitigation treatment.

11:15

1aNS13. Noise from rooftop mechanical equipment in residential buildings. Anthony Nash (Charles M. Salter Assoc., 130 Sutter St., Ste. 500, San Francisco, CA 94104, anthony.nash@cmsalter.com)

Community planning codes restrict building heights with the intent that the building is in general conformance with the character of the neighborhood's built environment. Although these codes prescribe the maximum height, they do permit a limited amount of rooftop space to be used for elevator penthouses, exhaust fans, cooling towers, and miscellaneous other rotating mechanical equipment. Thus, the stage is set for a potential acoustical conflict between the need for building mechanical services on the rooftop above and the expectations of people occupying residences below (ironically, the most expensive space in the building). This paper discusses a number of design and retrofit projects involving rooftop equipment installations over noise-sensitive spaces. The cases include rooftop equipment such as small corridor exhaust fans, rotary chillers, traction elevators, etc. The paper describes both the noise generating mechanisms and transmission paths as well as techniques for mitigating mechanical equipment noise into occupied spaces.

11:30

1aNS14. Active feedback control of broadband noise from a small axial fan. Cole V. Duke, Scott D. Sommerfeldt, and Kent L. Gee (Dept. of Phys. and Astron., Brigham Young Univ., N283 ESC, Provo, UT 84602)

In the past, significant progress has been made in actively controlling the tonal noise of axial cooling fans. Current work focuses on the control of broadband fan noise, which presents different challenges. Because the broadband noise is not deterministic, a feedback system will be used, and the total system delay must be reduced as much as possible. To ensure stability of the closed-loop control system, the phase and magnitude characteristics of the filter must be chosen to produce attenuation in the target frequency band without augmenting the noise outside the band. A practical analog feedback controller will be presented, and results will be shown for a single input single output control system, as well as a configuration using multiple independent control systems.

11:45

1aNS15. The prediction of airflow generated noise in ventilation system bends. David C. Waddington (Acoust. Res. Ctr., Univ. of Salford, Salford M5 4WT, UK, d.c.waddington@salford.ac.uk) and David J. Oldham (Univ. of Liverpool, Liverpool L69 3BX, UK)

This paper describes recent work on the prediction of airflow generated noise in ducts based upon pressure loss characteristics. Strategies employed to reduce turbulence by smoothing the flow around a bend include the use of turning vanes, and a large radius of curvature with the associated use of splitters. The effect of these strategies on the regenerated noise is examined for a range of common bends including mitred bends, with and without turning vanes, and radiussed bends, with and without splitters. The effect of bends with different aspect ratios is also examined. An excellent collapse of measured data onto a single characteristic curve is reported for each configuration based upon standard values of pressure loss coefficients. It is suggested that that these characteristic curves might form the basis for a practical prediction method.

Session 1aSCa

Speech Communication: Articulatory Speech Synthesis and Robotic Speech

Michael C. Brady, Cochair
Dept. of Cognitive Science, Univ. of Indiana, Bloomington, IN 47406

Frank H. Guenther, Cochair
Dept. of Cognitive and Neural Systems, Boston Univ., Boston, MA 02215

Invited Papers

8:30

1aSCa1. Three-dimensional mechanical tongue and vocal cord models for human speech production. Kotaro Fukui, Eiji Shintaku, Yuma Ishikawa, Nana Sakakibara, Yoshikazu Mukaeda, Atsuo Takanishi (Dept. of Modern Mech. Eng., School of Creative Sci. and Eng., Waseda Univ., 3-4-1, Ookubo, Shinjuku-ku Tokyo, 196-8555, Japan, kotaro@toki.waseda.jp), and Masaaki Honda (Waseda Univ., Saitama, Japan)

We have developed talking robot Waseda Talker series, which mimicked human speech production mechanism. This robot is consisted of mechanical vocal cords, tongue, lips, and nasal cavity. The current version, WT-7R (Waseda Talker No. 7 Refined), has 16 DOF (7 DOF in tongue, 1 DOF in jaw, 4 DOF in lips, 1 DOF in velum, 2 DOF in vocal cords, and 1 DOF in lung), having same vocal tract length as average adult male. The mechanical vocal cord model is made of styrene based thermoplastic elastomer "Septon" and the shape is mimicking vocal folds of human with three dimensions. The vocal code model has pitch control mechanism by adjusting the cord tension and glottal opening-closing mechanism, and it could reproduce the vocal cord vibration of the vocal fry (double pitched) and the breathy voice (unclose in one cycle) as well as the modal voice with variable pitch. The three-dimensional tongue model is constructed by the rigid link mechanism covered by the Septon soft rubber. The tongue model could be controlled so as to configure the tongue shape in a three-dimensional way. We will describe the details of the mechanism of talking robot by showing some demonstration videos.

8:50

1aSCa2. Gesture controlled synthetic speech and song. Sidney Fels (Elec. and Comput. Eng., Univ. of British Columbia, 2332 Main Mall, Vancouver, BC, Canada, V6T 1Z4, ssfels@ece.ubc.ca), Bob Pritchard (Univ. of British Columbia, Vancouver, BC, Canada V6T 1Z2), Eric Vatikiotis-Bateson (Univ of British Columbia, Vancouver BC, V6T 1Z4 Canada, and The Visual Voice Team)The Visual Voice Team

We describe progress on creating digital ventriloquized actors (DIVAs). DIVAs use hand gestures to synthesize audiovisual speech and song by means of an intermediate conversion of hand gestures to articulator (e.g., tongue, jaw, lip, and vocal chords) parameters of a computational three-dimensional vocal tract model. Our parallel-formant speech synthesizer is modified to fit within the MAX/MSP visual programming language. We added spatial sound and various voice excitation parameters in an easy-to-use environment suitable for musicians. The musician's gesture style is learned from examples. DIVAs will be used in three composed stage works of increasing complexity performed internationally, starting with one performer initially and culminating in three performers simultaneously using their natural voices as well as the hand-based synthesizer. Training performances will be used to study the processes associated with skill acquisition, the coordination of multiple "voices" within and among performers, and the intelligibility and realism of this new form of audio/visual speech production. We are also building a robotic face and computer graphics face that will be gesture controlled and synchronized with the speech and song. [This project is funded by the Canada Council for the Arts and Natural Sciences and Engineering Research Council, Canada. More information is at: www.magic.ubc.ca/VisualVoice.htm]

9:10

1aSCa3. Biomimetic vocal tract modeling: Synthesis of speech articulation. Robin Hofe and Roger K. Moore (Univ. of Sheffield, Dept. of Comp. Sci. Regent Court, 211 Portobello, S1 4DP, Sheffield, UK, R.Hofe@sheffield.ac.uk)

In recent years, a number of researchers have started to investigate cognitive aspects of speech production using "speaking robots." This paper introduces AnTon–the Animatronic Tongue and vocal tract model developed at the University of Sheffield, UK. AnTon's distinctive feature is its design principle. Whereas other speaking robots are motivated by functionality—to produce a good imitation of human speech—AnTon mimics human anatomy, including tissue properties and muscle (actuator) configurations. Any ability to produce speech sounds emerges solely as a consequence from that anatomy. Such imitation of a biological system is called "biomimetic." The reason to choose a biomimetic approach lies in the need for an experimental tool that makes speech articulation processes transparent. The use of realistic muscle configurations means that the model's speech gestures can be mapped directly to human behavior, facilitating the evaluation of both the quality of AnTon's speech gesture reproductions as well as any predictions about human behavior inferred from experiments with AnTon. AnTon will be used to clarify speech articulation processes, specifically in relation to energy efficiency of articulatory gestures and the importance of an auditory feedback loop. The project's progress is continually documented on http://www.dcs.shef.ac.uk/~robin/anton/anton.html.

9:30

1aSCa4. Synthesizing vowel transitions with an analog vocal tract. Michael C. Brady (Dept. of Cognit. Sci. Indiana Univ., 1910 E. 10th St., Bloomington, IN 47406, mcbrady@indiana.edu)

A mechanical vocal tract with a pneumatic sound source, silicone tongue, and lip rounding mechanism is introduced. The tract is designed to make controlled transitions between static articulatory vowel configurations. The focus on transitions is important because many argue that it is the change between steady-state sounds that the nervous system is tuned for in extracting information from the speech signal. I draw on examples and experimental results as I review this steady state versus transition distinction. The notion of articulatory transitions as the motor control targets of robotic speech production is then discussed. Video demonstrations of the mechanical tract helps to illustrate how transitions as control targets may be implemented. In conclusion, I argue for why an analog vocal tract with its true aerodynamics (as opposed to synthesis using a cascade of digital filters) is called for in generating articulator-transition-based speech sounds.

9:50

1aSCa5. A talking robot and the interactive speech training for hearing-impaired. Hideyuki Sawada (Dept. of Intelligent Mech. Syst. Eng., Faculty of Eng., Kagawa Univ., 2217-20 Hayashi-cho, Takamatsu, Kagawa, 761-0396, Japan, sawada@eng.kagawa-u.ac.jp)

A speech trainer employing a talking robot for auditory-impaired people will be introduced. The talking robot consists of mechanically-designed vocal organs such as a vocal tract, a nasal cavity, artificial vocal cords, and a sound analyzer with a microphone system, and the mechanical parts are fully controlled by ten servomotors for generating humanlike voices. The robot autonomously learns the relation between motor control parameters and the generated vocal sounds by an auditory feedback control, in which a Self-organizing neural network is employed for the adaptive learning. By employing the robot and its properties, the authors have constructed an interactive speech training system. The training is divided into two approaches; one is to use the robot for showing the shape and the motion of the vocal organs, and the other is to use a topological map for presenting the difference of phonetic features of a traine's voices. While referring to the robot motions and the phonetic characteristics, a trainee is able to interactively practice vocalization for acquiring clear speech with appropriate speech articulations. To assess the validity of the training system, a practical experiment was conducted by 19 subjects in a school for the deaf children. The talking robot is expected to intensively teach auditory-impaired patients the clear speech articulations.

10:10

1aSCa6. Real-time speech synthesis for neural prosthesis. Frank H. Guenther (Dept. of Cognit. & Neural Syst., Boston Univ., 677 Beacon St., Boston, MA 02215)

The idea of building an artificial vocal tract that could serve as a prosthetic device for mute individuals has been pursued since the pioneering work of Wolfgang von Kempelen and contemporaries in the 18th century. These early speaking machines were built out of wood, metal, rubber, and other materials fashioned into a mechanical vocal tract whose shape was manipulated manually by the user while air was forced through it. In the mid-20th century, Gunnar Fant and colleagues constructed electrical resonating circuits capable of producing speechlike sounds over loudspeakers, and shortly thereafter fully computer-based speech synthesizers were developed, including formant synthesizers and articulatory synthesizers. The recent development of brain-computer interfaces driven by neuronal activity raises the possibility of a speech prosthesis that would allow mute paralyzed individuals to directly control a speech synthesizer to create speech output. We have implemented such a brain-computer interface in a 25-year-old volunteer suffering from locked-in syndrome, characterized by complete loss of movement but intact perception and consciousness. To date the volunteer has successfully used the prosthesis to produce vowel sounds. Future developments in speech synthesis necessary for allowing full conversational speech via brain-computer interface will be outlined. [Work supported by NIH/NIDCD and NSF.]

Session 1aSCb

Speech Communication: Speech Production and Articulatory Models (Poster Session)

Michael C. Brady, Cochair
Dept. of Cognitive Science, Univ. of Indiana, Bloomington, IN 47406

Frank H. Guenther, Cochair
Dept. of Cognitive and Neural Systems, Boston Univ., Boston, MA 02215

Contributed Papers

All posters will be on display from 10:30 a.m. to 11:30 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 10:30 a.m. to 11:00 a.m. and contributors of even-numbered papers will be at their posters from 11:00 a.m. to 11:30 a.m.

1aSCb1. Response of synthetic vocal fold models with geometry based on visible human project data. Brian A. Pickup and Scott L. Thomson (Dept. of Mech. Eng., Brigham Young Univ., 435 CTB, Provo, UT 84602, thomson@byu.edu)

Numerous synthetic and computational models have been and are currently being used in research studies of human vocal fold vibration. Model geometry plays an integral role in governing dynamic response. However, the various model geometry definitions have typically been idealized and often exhibit wide variability with each other. The present research compares the response of synthetic vocal fold models of different geometries. One model is based on idealized geometry, while the other is based on geometry obtained from the National Library of Medicine's visible human project (VHP). The process for image extraction, model definition, and model fabrication is described, including: (1) image conversion from 2-D VHP image sequences to 3-D stereolithography (STL) format, (2) conversion to 3-D computer model format in which model geometric manipulations can be performed, and (3) fabrication of synthetic models using rapid prototyping. Results of measurements to characterize the dynamic response of self-oscillating synthetic vocal fold models, including onset and offset pressure, instantaneous glottal width using high-speed imaging, and glottal jet velocity profiles using particle image velocimetry (PIV), are presented for models based on both VHP data and idealized geometries. The sensitivity of the models to geometry changes is also reported.

1aSCb2. Analysis of recent experimental studies of phonation threshold pressure done with a physical model of the vocal fold mucosa. Lewis P. Fulcher (Dept. of Phys. and Astronomy, Bowling Green State Univ., Bowling Green, OH 43403), Ronald C. Scherer, Chen Liu, and Marco Nardone (Bowling Green State Univ., Bowling Green, OH 43402)

The pioneering work on the physics of small amplitude oscillations of the vocal folds by Titze resulted in a simple analytic formula that predicted the phonation threshold pressure of the uniform glottis to be a linear function of the glottal half-width, to be directly proportional to energy dissipation properties of the vocal fold, and to be reduced by the presence of a vocal tract. These relationships have been examined in a series of experiments [R. Chan and I. Titze, "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," J. Acoust. Soc. Am. **119**, 2351–2362 (2006)], and the linear connection with glottal half-width has been observed in several cases. However, the analysis described above did not consider the role of viscosity of the air, which may be important at small widths, as suggested by Lucero. The present study examines these experiments in the context of the surface wave model developed by Titze using intraglottal pressure distributions taken with the static physical model of the glottis M5. The M5 data include effects of viscosity and of flow separation. Preliminary results indicate that the threshold pressure should decrease as glottal width decreases, over the entire range of glottal widths examined by Chan and Titze. [Work supported by NIH R01DC03577.]

1aSCb3. Nonlinear viscoelastic response of the vocal fold lamina propria under large-amplitude oscillatory shear. Roger W. Chan (Otolaryngol., Biomed. Eng., Univ. of Texas Southwestern Med. Ctr., Dallas, TX 75390) and Miwako Kimura (Univ. of Texas Southwestern Med. Ctr., Dallas, TX 75390)

Viscoelastic properties of the vocal fold lamina propria have been reasonably quantified beyond the linear range for describing tissue behavior under uniaxial stretch or tensile deformation, but the same cannot be said for oscillatory shear deformation, which is more relevant for understanding the mechanics of large-amplitude vocal fold vibration. Previous studies reporting the viscoelastic shear properties of vocal fold tissues have been limited to characterization of the small-amplitude viscoelastic response in the linear viscoelastic region, partly due to the fact that derivation of the elastic and viscous shear moduli is based on the assumption of linearity, and also because of the lack of a coherent framework for describing any nonlinearities beyond the linear viscoelastic region. Based on a recently proposed rheological framework for quantifying such nonlinearities, this study examined the viscoelastic response of human vocal fold cover specimens subjected to large-amplitude oscillatory shear, up to a shear strain amplitude of around 1.0. Results indicated that the linear viscoelastic moduli (G' and G'') cannot adequately describe the tissue response under large-strain shear deformation, whereas geometrical interpretations of Lissajous–Bowditch curves could unveil nonlinearities that are obscured by the use of G' and G'', such as the phenomenon of strain stiffening.

1aSCb4. Effects of the airway surface liquid on vocal fold vibration. David Daily, Dan Lo Forte, and Scott L. Thomson (Dept. of Mech. Eng., Brigham Young Univ., 435 CTB, Provo, UT 84602)

A two-layer airway surface liquid (ASL), comprised of a Newtonian sublayer and a non-Newtonian surface layer, lines the laryngeal lumen. During phonation the layers on opposing vocal folds experience merging, squeezing, stretching, and rupture. Various aspects of the ASL have been studied, but its role in vocal fold vibration remains largely understudied. This presentation describes the results of experiments and computational simulations aimed at improving our understanding of the role of the ASL, in particular the non-Newtonian layer, in phonation. The experimental setup for measuring the liquid properties is described, and results are presented. The measured liquid properties are input into a finite element model of the vocal folds, which includes a simulated non-Newtonian layer, using the software package ADINA. The non-Newtonian liquid is also incorporated into a two-mass vocal fold model, which includes the effects of the ASL during colli-

sion and separation. The finite element and two-mass models are compared with each other and with experimental results to assess the influence of the non-Newtonian component of the ASL on vocal fold vibration.

1aSCb5. Rapid three-dimensional magnetic resonance imaging of vocal tract shaping using compressed sensing. Yoon-Chul Kim, Shrikanth Narayanan, and Krishna S. Nayak (Ming Hsieh Dept. of Elec. Eng., Univ. of Southern California, 3740 McClintock Ave., Los Angeles, CA 90089, yoonckim@usc.edu)

Conventional 3D magnetic resonance imaging (MRI) of vocal tract during sustained sound production takes prohibitively long and requires repetitions of the same articulation with pauses of the scan. This can produce image misregistration artifacts possibly due to different positioning of the jaw, tongue, and lips at each repetition. Within a scan time limit of 6–10 seconds, a duration practical for sustained sound production, it is desirable to accelerate high-resolution 3D imaging by undersampling 3D *k*-space. In this study, compressed sensing theory was adopted to achieve the acceleration by incoherent undersampling of 3D *k*-space and minimum L1-norm reconstruction. The increase in transform sparsity was exploited by constraining image phase in which rapid phase variation is expected in air-tissue boundaries due to large magnetic susceptibility difference. Vocal tract volume of interest was imaged during sustained sound production of English consonants /s/, /sh/, /r/, and /l/. Vocal tract area functions were measured using OsiriX software. The visualization of 3D tongue shape was demonstrated via image segmentation and volume rendering process. MR acquisition for each consonant took 6 sec without any repetitions of the articulation. The reconstruction of a whole 3D vocal tract was achievable with a 1.33 $\times$1.33$\times$1.33 mm^3 isotropic resolution. [Work supported by NIH.]

1aSCb6. An analysis-by-synthesis approach to modeling real-time MRI articulatory data using the task dynamic application framework. Erik Bresch (Dept. of Elec. Eng., Univ. of So. California, Los Angeles, CA 90089, bresch@usc.edu), Louis Goldstein, and Shrikanth Narayanan (Univ. of So. California, Los Angeles, CA 90089)

We report on a method of modeling real-time MRI articulatory speech data using the Haskins task dynamic application (TaDA) framework. TaDA models speech using a set of discrete dynamical regimes that control the formation of vocal tract constrictions (gestures). An utterance can be specified by a gestural score: the pattern of activation of these regimes in time. Individual model articulator degrees of freedom are automatically coordinated according the concurrent demands of the unfolding constrictions. Our modeling procedure consists of two stages: (1) After determining the outline of the midsagittal upper airway, time series of constriction measurements are derived which allow the estimation of the subject-specific parameters relating the articulator and constriction domains. (2) Gradient descent is utilized to adjust the activation intervals of the gestural score generated by TaDA for that utterance so that the synthesized vocal tract constriction evolution matches the observed MRI time series. Additionally, information related to glottal control, which cannot be derived from the MR images, is extracted from the simultaneously recorded audio, and it is used to optimize the timing of glottal abduction gestures in the score. Adequacy of fit is evaluated at articulatory, constriction, and perceptual levels. [Work supported by NIH.]

1aSCb7. Vowel recognition from articulatory position time-series data. Jun Wang (Dept. of Comput. Sci. & Engr., Univ. of Nebraska-Lincoln, Lincoln, NE 68508, junwang@cse.unl.edu), Ashok Samal , Jordan Green , and Tom Carrell (Univ. of Nebraska-Lincoln, Lincoln, NE 68583)

The goal of this project is to determine the accuracy and processing speed of different approaches for mapping time-varying articulatory positional data to vowels. Three widely used classifiers were compared on two datasets: one single speaker and one multiple speaker. The single-speaker dataset was acquired using the Articulograph AG500. The multiple-speaker dataset was obtained from seven speakers in the Xray Microbeam Speech Production Database (Westbury, 1994). The recognition rate for single speaker dataset (eight English vowels) ranged from 94.25% to 98.1%, and from 62.38% to 99.35% for the multiple-speaker dataset. For the single-speaker dataset, recognition accuracy was comparable across classifiers. For the multiple-speaker dataset, recognition accuracy was better for the Support Vector Machine and C4.5 than for the neural networks. The decision tree generated by C4.5 was consistent with articulatory features commonly used to descriptively distinguish vowels. Moreover, the Support Vector Machine and C4.5 performed much faster than did the neural network. The high recognition rates observed suggest that static vowels can be accurately recognized from articulatory position time-series data. The findings are intended to improve the accuracy and response time of a real-time articulatory-movement based speech synthesizer.

1aSCb8. Recovering area functions from acoustics and midsagittal profiles. Richard S. McGowan and Rebecca W. McGowan (CReSS LLC, 1 Seaborn Pl. Lexington, MA 02420)

Simultaneously recorded midsagittal pellet positions and acoustic output for several subjects comprise the Wisconsin X-Ray Microbeam Speech Production Database (XRMB-SPD). An articulatory synthesizer, EASY, was used to help recover area function parameters corresponding to the production of vowels by subjects in the XRMB-SPD by means of analysis-by-synthesis. Specifically, for each subject, an initial personalized midsagittal profile was created based on the pellet positions, pellet trace, and pharyngeal wall trace during neutral vowel production. Based on the scales of this initial profile for a subject, subsequent midsagittal profiles were generated from the pellet data during various vowel productions. Each midsagittal profile and the corresponding first three formant frequencies were used to optimize the parameters of an area function in the EASY for closest acoustic match. The optimization was performed with a genetic algorithm. The results are examined in terms of consistency within a subject for different vowels and consistency within vowel for different subjects. Synthesized examples of vowel production will also be presented. [Work partially supported by Grant DC-001247 to CReSS LLC.]

1aSCb9. Korean articulatory speech synthesis using physical vocal tract model. Huynh V. Luong (Univ. of Ulsan, Korea), Jong-Myon Kim (Univ. of Ulsan, Korea), and Cheol Hong Kim (Chonnam Natl. Univ., Korea)

Artificial vocal tract models provide the support of learning a second language and the therapy of speech disorders. Moreover, phonetic education and research can benefit from articulatory speech synthesis. Articulatory speech synthesis models are constructed by the source-filter model of the human vocal tract. In this study, we generated a Korean articulatory speech synthesis model using Artisynth [Fels *et al.*, ISSP, 419–426 (2006)], which is a 3-D biomechanical open-source simulation platform. As the origin of the Korean language, it has 10 basic vowel phonemes and 11 complicated vowels in which some vowels can be rounded and unrounded such as /eu/, /yeo/, /wae/, etc. To synthesize these specific vowels, we created a new physical vocal tract model, which interconnects to form a complete integrated biomechanical system. The created model efficiently supports recording the Korean vowel sounds and linguistic analysis based on the linear prediction model. As a result, parameters of the glottis and controllable vocal tract filter are automatically evaluated. The acoustic quality of the synthesizer for Korean vowels is comparable with that of the existing commercial speech synthesis systems such as concatenation synthesizers [Donovan (1996)] and [Hamza (2000)]. [Work supported by the MKE, Korea, under the ITRC supervised by IITA (IITA-2008-(C1090-0801-0039)).]

1aSCb10. Vocal and postural coupling in singing and speech. Eric Vatikiotis-Bateson, Martin Oberg, Adriano Barbosa (Linguist., Univ. of British Columbia, 2613 West Mall, Vancouver, BC, Canada, V6T 1Z4), Nancy Hermiston, and Richard Kurth (Univ. of British Columbia, Vancouver, BC, Canada, V6T 1Z1)

This study examines behavioral measures during speech and singing to assess the interaction of the postural control system and the production of expressive vocal behavior at different levels of vocal effort. We know that head motion is essential to both expressive speech and to balanced posture, suggesting that the head must be entrained to some extent by both subsystems simultaneously. This study shows that increased vocal effort (defined here as loudness) affects rigid body (6D) head motion and other measures of postural behavior acquired from force plates (6D) and video-

based motion analysis (2D). Specifically, as vocalizations becomes louder, the postural system becomes more simply coupled spatially and temporally to the vocal system.

1aSCb11. Optimization of articulator trajectories in producing learned nonsense words. Mark Tiede (Haskins Lab., 300 George St., New Haven, CT 06511 & Speech Comm. Group, MIT-R.L.E., tiede@haskins.yale.edu), Christine Mooshammer (Haskins Laborities, New Haven, CT 06511), Louis Goldstein (U.S.C., Los Angeles, CA 90089), Stefanie Shattuck-Hufnagel, Joseph Perkell, and Melanie Matthies (MIT-R.L.E., Cambridge, MA 02139)

Previous work assessing motor learning in novel speech utterances has shown that kinematic duration and variability are reduced as a function of practice [Schulz *et al.* (2001)]. In this study EMA was used to observe articulator movements during production of a variety of polysyllabic nonsense words conforming to English phonotactics (e.g., "thraimpoframodis"). Target words were elicited eight times in block-random order, with each block separated by ten minutes of an unrelated task serving as a distractor. Analysis of EMA sensor trajectories between initial and final repetitions shows in general reduction of overall duration, lower STI [Smith *et al.* (1995)], reduced stiffness, and fewer acceleration peaks, with intermediate productions indicating that these trends occur asymptotically. Comparison of consonant closure timings delimited using velocity extrema and aligned through nonlinear time normalization [Lucero *et al.* (1997)] suggests that as fluency increases, the overlap between adjacent consonantal gestures also increases. The patterns of articulatory optimization illustrated by these results will be contrasted with similar analyses of real but difficult-to-pronounce sequences (tongue twisters) collected from the same subjects, with the goal of identifying similarities between pre-fluent and dysfluent utterances. [Research supported by NIH.]

1aSCb12. Measuring articulatory similarity with algorithmically reweighted principal component analysis. Jeff Mielke and Joseph Roy (Arts 401, 70 Laurier Ave. E., Ottawa, ON Canada, K1N6N5, jmielke @uottawa.ca)

Articulatory similarity was assessed for a corpus of 2700 ultrasound images of 75 cross-linguistically frequent speech sounds produced by four subjects in three segmental contexts (a_a, i_i, u_u). Cross-distances were generated for the entire length of the vocal tract using the Palatron algorithm and realistic estimates of the location of the pharyngeal wall and teeth, resulting, after interpolation, in 60 cross-distances per token. A standard principal component analysis of these data is overwhelmed by the coarticulatory effects of the context vowels. Algorithmically reweighted principal component analysis was devised in order to use coarticulatory variation to isolate the most distinctive cross-distances for each target segment. The reweighting algorithm considers the variance across repetitions in each of the 60 cross-distances, as well as variance in the slope of the cross-distance function, in order to identify areas of stability across tokens. For each target sound, cross-distances with the greatest variance are reset to the mean for all target segments, while cross-distances with the least variance maintain their original values. This has the effect of defining each target sound by the cross-distances, which are most stable across contexts.

1aSCb13. The influence of dynamic flow conditions on sound generation in fricative consonants. Gordon Ramsay (Haskins Lab., 300 George St., New Haven, CT 06511)

Studies of speech production have often remarked that frication noise in voiced fricatives is weaker than that produced in homorganic voiceless fricatives. This is traditionally attributed to the reduction in mean flow (and, hence, Reynolds Number) produced by the glottal constriction needed to induce voicing. However, research in fluid mechanics has demonstrated that turbulence may develop quite differently in steady and unsteady background flows. Strong flow gradients may inhibit the transition to turbulence, or even relaminarize an already-developed turbulent flow. This paper describes a series of mechanical modelling experiments designed to investigate the effect of dynamic flow conditions on sound generation in fricatives. A feedback-regulated Kolvenbach flow valve, controlled by computer, is used to create carefully-shaped dynamic modulations of the air flow through a stereolithographic replica of the vocal tract, about a series of static flow conditions. A multitaper spectral analysis technique is then used in conjunction with sweeptone measurements to separate out the contribution of source and filter components to the far-field sound spectra recorded with a microphone, and source properties are compared across flow conditions. The observed variations cannot be explained by differences in mean flow alone, suggesting that dynamic effects may be important.

1aSCb14. Effects of sleep deprivation on nasalization in speech. Xinhui Zhou (Dept. of Elect. and Comp. Eng., Univ. of Maryland, College Park, MD 20742, zxinhui@glue.umd.edu), Suzanne Boyce (Univ. of Cincinnati, Cincinnati, OH 45267), Joel MacAuslan (Speech Technol. and Appl. Res. Corp., Bedford, MA 01730), Walter Carr, Balkin Thomas, Picchioni Dante (Walter Reed Army Inst. of Medicine, Silver Spring, MD 20910), and Carol Espy-Wilson (Univ. of Maryland, College Park, MD, 20742)

Various aspects of speech appear to be subtly affected by sleep deprivation [Boyce *et al.*, "Landmark-based analysis of sleep-deprived speech," J. Acoust. Soc. Am. **123**(5), 3887 (2008)]. An automatic procedure for detection of nasalization in the acoustic spectrum of vowels was applied to a database of speech recordings from 15 subjects in rested and sleep-deprived conditions. Results showed acoustic differences consistent with differences in sleep condition, suggesting that lack of sleep affects the motor organization/execution of speech articulation in general, and for oral/nasal coupling during vowels in particular. Significant differences in male vs female patterns were also observed. Correspondences between these patterns and those reported in the clinical and linguistic literature will be discussed. [Work supported by NIH.]

1aSCb15. Gestural intrusions in Taiwanese and English. Man Gao (Haskins Labs., 300 George St., Suite 900, New Haven, CT 06511), Yueh-chin Chang, Feng-fan Hsieh (Grad. Inst. of Linguist., Natl. Tsing Hua Univ., Taiwan), Hosung Nam, Mark Tiede (Haskins Labs., New Haven, CT 06511), and Louis Goldstein (U.S.C, Los Angeles, CA 90089; & Haskins Labs., New Haven, CT 06511)

Kinematic studies of repetitive speech production tasks involving English sequences such as "cop top" have provided abundant evidence for gestural intrusion errors: errorful constrictions that are co-produced with intended target constrictions (Pouplier and Goldstein, 2005; Goldstein *et al.*, 2007). These intrusions have been analyzed as transitions to more dynamically stable coordination modes. This work investigates Taiwanese speakers' production of similar repetitive sequences such as [kap tap]. If the speech sequences are planned and executed in Taiwanese in the same way as in English, similar patterns of intrusive speech errors should be observed. However, analysis of the kinematic data from three speakers shows that these effects in Taiwanese differ from those in English in a number of ways. A systematic comparison of Taiwanese and English intrusions is presented in terms of frequency of occurrence (much higher in Taiwanese), standard deviation of magnitude (smaller in Taiwanese), and temporal relationships to intended gestures. The results are discussed with respect to the language-specific regularities of gestural coordination and their implications for the mode-locking account of speech errors. [Work supported by NIH-NIDCD DC008780.]

Session 1aUW

Underwater Acoustics: Propagation Modeling

Eric I. Thorsos, Chair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105-6606

Chair's Introduction—7:55

Contributed Papers

8:00

1aUW1. Transport theory for shallow water propagation with rough boundaries. Eric I. Thorsos, Frank S. Henyey, W. T. Elam, T. Brian Hefner, Stephen A. Reynolds, and Jie Yang (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

At frequencies of about 1 kHz and higher, forward scattering from a rough sea surface (and/or a rough bottom) can strongly affect shallow water propagation and reverberation. The need exists for a fast, yet accurate method for modeling such propagation where multiple forward scattering occurs. A transport theory method based on mode coupling will be described that yields the first and second moments of the field. This approach shows promise for accurately treating multiple forward scattering in one-way propagation. The method is presently formulated in two space dimensions, and Monte Carlo rough surface PE simulations are used for assessing the accuracy of transport theory results. [Work supported by ONR Ocean Acoustics.]

8:15

1aUW2. Wave refraction at an interface: Snell's law versus Chapman's law. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA/Earth System Res. Lab., Boulder, CO 80305, Oleg.Godin@noaa.gov)

Wave energy streamlines determine the paths along which energy is transported from a wave source. The energy streamlines provide insights into mechanisms of wave propagation and scattering and are often utilized to visualize wave fields. In contrast to rays, which are essentially an asymptotic, short-wave concept, energy streamlines adequately represent arbitrary wave fields. However, usefulness of energy streamlines in studies of wave fields is limited by the fact that, unlike rays, no general laws governing energy streamline refraction are known. Here a simple refraction law is derived for energy streamlines of acoustic and linearly polarized electromagnetic waves. The refraction law is an extension of a result recently established by D. M. F. Chapman for sound refraction at plane interfaces of homogeneous media. While Snell's law of ray refraction at an interface is formulated in terms of the ratio of sound speeds in media in contact in the case of acoustic waves, Chapman's law implies that refraction of acoustic energy streamlines is controlled by the ratio of mass densities. Similarities and differences between Snell's and Chapman's refraction laws are discussed. It is shown that analysis of energy streamlines provides a new, helpful perspective on wave transmission through an interface.

8:30

1aUW3. Parabolic equation modeling of anisotropic waves in heterogeneous media. Adam M. Metzler, William L. Siegmann (Rensselaer Polytech. Inst., Troy, NY 12180), Ralph N. Baer, and Michael D. Collins (Naval Res. Lab., Washington, DC 20375)

Wave propagation in heterogeneous media can be modeled efficiently with the parabolic equation method, which has been extended to problems with homogeneous anisotropic layers [A. J. Fredricks *et al.*, Wave Motion **31**, 139–146 (2000)]. This approach is currently being extended to the heterogeneous case, including piecewise continuous vertical dependence and horizontal dependence that is relatively gradual but which may be large over sufficient distances. Vertical dependence is included by applying appropriate heterogeneous depth operators in the equations of motion. Horizontal dependence is included by extending the single-scattering solution [E. T. Kusel *et al.*, J. Acoust. Soc. Am. **121**, 808–813 (2007)] to the anisotropic case. [Work supported by the Office of Naval Research.]

8:45

1aUW4. A two-dimensional coupled mode model for acoustic propagation in a range-dependent waveguide. Wenyu Luo and Henrik Schmidt (Dept. of Mech. Eng., MIT, 77 Massachusetts Ave., Cambridge, MA 02139, lwy@mit.edu)

A two-dimensional, two-way coupled mode model is presented in this work. By combining a forward marching and a backward marching, this model is valid for problems where backscattering is significant. This two-way solution provides great accuracy, and it also features in efficiency, which is achieved by applying the single-scatter approximation. In addition, this model is stable with appropriately normalized range solutions in the formulation. This model is capable of handling two-dimensional problems with either a point source in cylindrical geometry or a line source in plane geometry. The computer code CSNAP is adopted to provide the fundamental model solutions and their associate coupling matrices in this model. This two-dimensional model can be easily extended to a three-dimensional model with the feasibility to solve large-scale realistic three-dimensional problems. Numerical examples are presented to demonstrate the accuracy, efficiency, and stability of this two-dimensional, two-way coupled mode model. [Work supported by the Office of Naval Research.]

9:00

1aUW5. Shallow water tomography: Ray theory vs travel-time sensitivity kernels. I. Iturbe (GIPSA-lab, 961 Rue de la Houille Blanche, BP 46, 38402 Saint Martin d'Heres, France, ion.iturbe@gipsa-lab.inpg.fr), P. Roux (Univ. Joseph Fourier, 38042 Grenoble, France), B. Nicolas (GIPSA-lab, 38402 Saint Martin d'Heres, France), J. Virieux (Univ. Joseph Fourier, 38042 Grenoble, France), and J. Mars (GIPSA-lab, 38402 Saint Martin d'Heres, France)

In underwater acoustics, ray tomography is the classical method used to estimate velocity variations, but recently. Travel-Time Sensitivity Kernels (TSK) approaches have been developed. In this paper, we deal with TSK for two source-receive arrays in an acoustic waveguide for shallow water tomography. As a first step, we show that separation of the different raypaths is improved by using a recently proposed new array processing [time-delay double beamforming (DBF) algorithm]. DBF consists of changing the 3-D data space from source depth, receiver depth and time into a new 3-D space related to ray propagation expressed by the beamformed variables, source angle, receive angle and time. As a consequence, each eigenray of the multipath propagation for a source-receiver couple can be identified and separated through DBF. In this context of DBF, the TSK is no longer point-to-point as usual, but relies on all source-receiver time series. Kernels are computed using the fact that the processed signal is a linear combination of time-delayed signals between all sources and receivers. Results in simulated data and in real datasets recorded in an ultrasonic tank prove that combination of TSK and DBF increases the resolution and robustness performance of shallow-water acoustic tomography.

9:15

1aUW6. Computationally efficient parabolic equation solutions to seismoacoustic problems. Jon M. Collis (Math and Comput. Sci., Colorado Sch. of Mines, 1500 Illinois St., Golden, CO 80401, jcollis@mines.edu)

Improving the capabilities and accuracy of the elastic parabolic equation method has been an active area of research. Variable rotation [D. A. Outing *et al.*, J. Acoust. Soc. Am. **120**, 3534–3538 (2006)] is an effective solution for treating variable bathymetry and has recently been extended to handle variable sediment thicknesses [J. M. Collis *et al.*, J. Acoust. Soc. Am. **123**, 51–55 (2008)] and beach interactions. To accurately resolve the field in elastic sediments, much finer grid spacing may be needed than is necessary to propagate the acoustic field in the water. In particular, as the shear wave speed in the elastic medium approaches zero, the governing system becomes singular, and very fine grid spacing is necessary to capture subtle effects of the field. In this presentation we show how elastic parabolic equation solutions can be computed with greater efficiency, through the use of multiple resolution grids. Interface conditions and derivative approximations on a nonuniform grid will be described. Finally, applications of multiple grid solutions will be discussed for rough surface scattering and treating near-singular low-shear speed sediments.

9:30

1aUW7. Acoustic wave scattering in turbulence and in internal waves. Tokuo Yamamoto, Clair Paris (AMP, RSMAS, Univ. of Miami, 4600 Rickenbacker Cswy, Miami, FL 33149, tyamamoto@rsmas.miami.edu), and Mohsen Badiey (Univ. of Delaware, Newark, DE 19716)

Analytical solutions of acoustic wave scattering have been obtained for turbulent currents and for shallow water internal waves. The scattering characteristics of the two are quit different. The most striking difference is that the maximum scattering is in only one direction by turbulence. Whereas there are two directions of maximum scattering in internal waves. In turbulence, acoustic waves scatter quite strongly while internal waves acoustic waves scatter very weakly. Therefore, turbulence changes acoustic wave (coherent) energy rapidly into (incoherent) scattered acoustic energy. The scattering of acoustic energy by internal waves is weak, on the order of 0.001 of that of turbulence. The theoretical results are verified with transmission loss measured in Kanmon Strait and at New Jersey Shelf with good agreements. [Work supported by ONR Code 322 OA.]

9:45

1aUW8. Resonant forward scattering of sound in the deep ocean by internal tides. Irina I. Rypina (Dept. of Physical Oceanogr., Woods Hole Oceanograph. Inst., 266 Woods Hole Rd., Woods Hole, MA 02543), Ilya A. Udovydchenkov, Timothy F. Duda (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), and Michael G. Brown (Univ. of Miami, Miami, FL 33149)

The importance of resonant forward scattering of sound by highly directional large amplitude nonlinear internal waves in shallow water environments has been well established. In contrast, most deep ocean environments are characterized by the presence of a sound speed perturbation (due mostly to internal waves) whose horizontal wave number spectrum is broad and approximately horizontally isotropic; under such conditions, the importance of resonant forward scattering is less clear. Exceptional deep ocean conditions are found in the vicinity of submarine ridges, which serve to generate internal tides that are both highly directional and have a narrow horizontal wave number spectrum. In this talk, ray, mode, and PE simulations are presented to explore the importance of resonant forward scattering by such structures under fairly realistic oceanographic conditions. [Work supported by ONR.]

10:00—10:15 Break

10:15

1aUW9. The second order resolution operator of a nonlinear ocean acoustics inverse problem. Robert I. Odom and Andrew A. Ganse (Appl. Phys. Lab, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, odom@apl.washington.edu)

The resolution operator for a linear inverse problem indicates how much smearing exists in the map between the true model and the estimated model. The trace of the resolution operator provides an estimate of the number of model parameters model that are resolved. In a series representation of the resolution operator for a nonlinear problem, the higher-order terms indicates how much spurious nonlinear leakage there is from the true model to the estimated model. In previous work, the solution of a simple nonlinear ocean acoustic inverse problem as a perturbation series in the horizontal wavenumber was constructed, and the linear data kernels were presented. This linear problem permitted the quantification of the magnitude of the perturbation and indicated when nonlinear effects must be taken into consideration. In this extension of previous work, the second order resolution kernel is constructed, which describes how effectively nonlinear effects can be removed from the reconstructed model. [Work supported by ONR.]

10:30

1aUW10. A finite element model for acoustic propagation in shallow water waveguides. Marcia Isakson (Appl. Res. Lab., The Univ. of Texas at Austin, Austin, TX 78713, misakson@arlut.utexas.edu)

A finite element model is developed using a commercial code to calculate the backscattering of acoustic waves in a two-dimensional shallow water waveguide with rough interfaces. Finite elements approach an exact solution to the Helmholtz equation as the discretization density increases. A time series for the backscattering is computed from time harmonic computations using Fourier synthesis. A parametric study of the roughness parameters for the sediment/water interface and the air/water interface will be presented in order to assess the range of validity of approximate methods. [Work sponsored by Internal Research and Development, Applied Research Laboratories.]

10:45

1aUW11. A comparison of acoustic uncertainty approximations in a Pekeris waveguide. David R. Dowling and Kevin R. James (Dept. of Mech. Eng., Univ. of Michigan, 2212 G. G. Brown, Ann Arbor, MI 48109, drd@engin.umich.edu)

Uncertainty in environmental parameters is often the dominant source of error in underwater acoustic field predictions. This presentation provides comparisons of three techniques for assessing uncertainty in a predicted acoustic field caused by uncertainty in environmental parameters: field shifting, polynomial chaos expansion, and coarse uniformly sampled direct simulations. The uncertainty assessments are performed in a 100-m deep Pekeris waveguide with an uncertain water-column sound speed for frequencies of 100 Hz to 1 kHz at ranges of 1 to 10 km with a variety of common bottom types. The accuracy and computational efficiency of each approximation are quantified in terms of an absolute-difference error norm for the probability density function (PDF) of acoustic field amplitude and the number of acoustic field calculations necessary to predict this PDF, respectively. In all cases, the true field-amplitude PDF is determined from numerically converged direct numerical simulations. The strengths and limitations of each technique are highlighted and the applicability of these results to other sound channels and uncertain parameters is discussed. In particular, polynomial chaos expansion produces the highest accuracy but does so at the highest computational cost, while field shifting produces acceptable accuracy for the least computational cost. [Work sponsored by the Office of Naval Research.]

11:00

1aUW12. Modeling gradients in ocean acoustics. Robert I. Odom (Appl. Phys. Lab, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, odom@apl.washington.edu) and Karsten James (California Polytechnic State Univ., San Luis Obispo, CA 93407)

Modeling acoustic propagation in the ocean requires a representation of the ocean sound speed. For a 1-D, range independent ocean, probably the most common representation comprises constant sound speed layers and/or gradient layers. Gradients of the form $1/c^2(z)$, admitting an exact solution for the acoustic pressure field in terms of Airy functions, are commonly employed. A disadvantage of using the Airy functions is that they are spe-

cific to that profile. For more general profiles, the best representation may be a stack of constant sound speed layers for which the pressure field is expressible in terms of simple exponentials. As an example, computing the backscattering strength from a stack of layers containing volume heterogeneities requires evaluation of an integral proportional to the 4th power of the pressure field. Integrating the simple exponential solutions for constant sound speed layers is trivial. Integrating a product of four Airy functions, either analytically or numerically, is not. If the medium is represented only by constant sound sound speed layers, what is the best way to choose the layers, and how many layers are needed? These questions are investigated by employing an algorithm originally due to Fuchs (1968). [Work partially supported by ONR.]

MONDAY AFTERNOON, 18 MAY 2009 GRAND BALLROOM II, 1:00 TO 4:50 P.M.

Session 1pAA

Architectural Acoustics and Noise: Acoustics of Green Buildings: A 360 Degree Perspective

David M. Sykes, Cochair
The Remington Group LP, 23 Buckingham St., Cambridge, MA 02138

Brandon D. Tinianov, Cochair
Serious Materials, 1250 Elko Dr., Sunnyvale, CA 94089-2213

Chair's Introduction—1:00

Invited Papers

1:05

1pAA1. National Academy of Engineering "Technology for a quieter America" project—impact on "green acoustics"? Patricia Davies (Sch. of Mech. Eng., Ray W. Herrick Labs., Purdue Univ., 140 Martin Jischke Dr., West Lafayette, IN 47907-2031, daviesp@ecn.purdue.edu) George C. Maling, Jr. (Inst. of Noise Control Eng., Harpswell, ME 04079) and Gregory C. Tocci (Cavanaugh Tocci Assoc., Inc., Sudbury, MA 01776)

In 2006, the National Academy of Engineering (NAE) initiated an investigation of the potential benefits of a focused effort to control noise in the workplace, communities, and homes that may be of interest to professionals engaged in "green" acoustics. It covered three initiatives: (1) applications of current technology; (2) research and development initiatives for noise control technology; and (3) intragovernmental and public relations programs. Several workshops were held in the 2005–2008 period with representatives from industry, academia, professional societies, private consultants, and government agencies who examined existing and potential solutions along with policies that would help to develop and deploy solutions. Members of the committee were George C. Maling, Jr., Robert J. Bernhard, Robert Bruce, Elizabeth A. Cooper, Patricia Davies, Carl Hanson, Robert Hellweg, Gerald Lauchle, Richard Lyon, Gregory Tocci, Ian Waitz, and others. In mid-2009, a consensus report will be issued that: (a) summarizes the state of current noise control engineering practice; (b) recommends how existing knowledge can address current challenges; (c) presents a research and education agenda that promotes the generation of new knowledge that can provide the greatest benefit to society; and (d) recommends policies that government agencies can develop to further improve the American soundscape.

1:20

1pAA2. LEED (Leadership in Energy and Environmental Design) and indoor environmental quality—What about office acoustics? Kevin M. Powell (Office of Appl. Sci. GSA Public Bldg.s Service, 555 Battery St., Rm. 518, San Francisco, CA 94111, kevin.powell@gsa.gov)

Recent research conducted by the US General Services Administration and the Center for the Built Environment disclosed an anomaly about "green building" interiors, that is, an apparently unintended consequence of the current LEED rating system. The anomaly is this: while occupants of offices that have been upgraded to achieve LEED certification express high levels of satisfaction with the "indoor environmental quality" of these spaces on most dimensions, there is one exception where scores actually decrease. This exception is the only dimension where occupant DISsatisfaction grows—it is acoustics. What is it about LEED-rated offices that cause this anomaly to occur? Why are occupants dissatisfied with the "acoustical comfort" of the new LEED-certified offices? Is it because there are more reflective, hard surfaces causing noise to reverberate? Or because the HVAC system has become quieter? Or because partition heights have been lowered to allow better airflow and more natural light, resulting in less privacy? Since increased occupant dissatisfaction may result in reduced productivity or increased absenteeism, the GSA's office of research has been examining this issue and will report on its findings and conclusions.

1:35

1pAA3. An acoustical consultant's experiences with the leadership in energy and environmental design (LEED) green building rating system(s). Ethan Salter (Charles M. Salter Assoc., Inc., 130 Sutter St., Ste. 500, San Francisco, CA 94104)

The panelist will discuss his experiences with several leadership in energy and environmental design (LEED) green building rating systems, developed by the U.S. Green Building Council (USGBC). He will provide insight into the process, advantages, and challenges of becoming a LEED accredited professional, his observations of how the promulgation of LEED acoustics credits will impact acoustics professionals, and also how the systems affect the solutions and products he recommends. The panelist became LEED accredited in 2006 and has consulted on several LEED projects across the country. He is active in his local chapter of the USGBC. Additionally, the panelist has attended the last four Greenbuild conferences put on by the USGBC.

1:50

1pAA4. The Human Factor: Healthcare, sustainability, and acoustics. Jane Rohde (JSR Assoc., Inc., 8191 Main St., Ellicott City, MD 21043, jane@jsrassociates.net)

There are several areas of concern within healthcare environments, which include staff turnover, infection control, patient satisfaction, and constant increase in costs with diminishing reimbursements. The healthcare systems used are under tremendous scrutiny and those that understand the necessity of creating sustainable care models are headed for future success. Sustainable care models are supported by the physical environment, which has to be based upon evidence-based design principles. This presentation will discuss a case study that is seeking to demonstrate the connection between the person, the place, and the product, resulting in improved patient and resident outcomes. Acoustics, both in the general population and the neonatal/pediatric population, is addressed. Examples of how noise causes disruption to sleep and negative patient and resident outcomes will be provided. Higher stress levels and other issues related to noise will be discussed regarding impacts on staff and performance. The sense of place and how this information can be impacted by design will then be reviewed. Many design professionals know little about acoustics or how to handle mitigating noise in the physical environment. The discussion will then turn to the discussion of products and how products tie into potential multi-faceted solutions.

2:05

1pAA5. Balancing acoustical performance with sustainability in LEED 2009 (Leadership in Energy and Environmental Design). Marsha Bischel and Amy Costello (Armstrong World Industries, Inc., 2500 Columbia Ave., Lancaster, PA 17603)

USGBC's LEED 2009 rating system provides minimal guidance for achieving optimal acoustical performance in sustainable spaces; however, there are opportunities within the rating systems to combine acoustics with other performance attributes to create spaces with optimal indoor environmental properties. Many acoustical finishes are composed of materials that can help achieve LEED points in several categories and can be critical differentiators in determining total building performance and occupant comfort. Strategies to be discussed will include highly light reflectant surfaces, high recycled content, passive energy reduction, and others. By employing such strategies, the acoustical performance of the building may be enhanced while meeting the sustainability objectives.

2:20

1pAA6. Application of acoustic materials for comfort and safety. Robert Matheson, Joseph King, and Natalia Levit (DuPont, 5401 Jefferson Davis Hwy., Richmond, VA 23234, natalia.v.levit@usa.dupont.com)

This work presents an overview of two high performance acoustic products from DuPont™ with various case studies on lowering low frequency noise transmission using DuPont™ LoWave™, low-weight thin blanket with enhanced low frequency performance, and lowering noise with cleanable, dust free sound absorptive panels protected by DuPont™ Tyvek®, a distinctive flash spun plexifilamentary film-fibril sheet, combining excellent barrier properties and characteristic porous structure to make it acoustically transparent in the voice frequency range.

2:35

1pAA7. Case studies and installation data on the acoustical properties of a new class of translucent, lightweight insulation material from NASA called aerogel. James Satterwhite (Cabot Corp., 157 Concord Rd., Billerica, MA 01821, james_satterwhite @cabot-corp.com)

Aerogels are well-known as a class of thermal insulation with "green" benefits including translucence, hydrophobicity, and light weight. However, before 2008, the acoustical properties of aerogels had not been characterized. Used by NASA in the 1970s, aerogels became available in the 1990s as insulation for the construction industry in skylights, underwater pipelining, and roofing fabric. In 2008, laboratory testing and field research began on the acoustical properties of thin profile (2–8 mm) architectural "tensile membrane" fabrics incorporating silica aerogel granules. Data from a tension structure in Vancouver—where aerogel-enhanced fabric was used to block aircraft noise—exhibited excellent acoustic absorption and acoustic impedance matching properties compared to insulators of comparable thickness. In this installation the material increased transmission loss of exterior to interior noise, and reduced indoor reverberation. Also in 2008, U.S. field tests demonstrated an aerogel blanket material as a surface treatment in offices to reduce broadband reverberation, resulting in increased speech intelligibility and enhanced acoustical comfort (an important factor in the 2009 LEED rating system). The acoustical attributes—combined with aerogel's thermal value, thin form factor, translucence, hydrophobicity, light weight, and absence of VOCs has led to growing interest in applications ranging from aircraft interiors to hospitals.

2:50—3:05 Break

Contributed Papers

3:05

1pAA8. Sustainablity and noise: An exploration of perception of noise as a significant component in sustainable design in a variety of cities internationally. Andreas Gerlsbeck (Zimmer, Gunsel, Frasca Architects, LLP., 320 SW Oak St., Rm. 500., Portland, OR 97204, agerlsbeck@zgf.com) and Paul van Orden (City of Portland Oregon, Portland, OR 97201)

The current acceptance in the sustainability design field of noise and acoustics as a holistically related design concern to the overarching discipline varies widely around the world. The authors' unique experiences on different continents help in the exploration of the current state of the acceptance of acoustical engineering in the sustainability field. An exploration of the paper from F. D. Henry [F. D. Henry, *Proc. 8th International Congress on Noise as a Public Health Problem* (2003)] is taken into account. The authors place notable focus on review of the efforts on the topic by several German research institutions such as: The German Federal Institute for Environmental Issues; The Federal Institute for Occupational Safety and Health; and The DEGA (The German Association of Acoustics). The authors expound on possible reasons why noise and acoustical design elements are often an afterthought by government development and environmental agencies, and design firms focused on sustainable design. The authors explore a number of suggestions for enabling forward movement toward increased inclusion of acoustical design in the sustainable design portions of future projects internationally.

3:20

1pAA9. Green building acoustics: Scoring user performance factors. Glenn E. Sweitzer (Sweitzer LLP, 4504 N. Hereford Dr., Muncie, IN 47304-1235, glenn.sweitzer@gmail.com)

Architectural acoustics lacks specificity in the LEED for Schools (2007), green building rating system, which is based on a checklist score. Uncorrected, this edition lacks the guidance needed to meet, consistently, the stated intent: to facilitate better teacher-to-student and student-to-student communications. An alternative system—user performance based—is developed for scoring speech transmission index (STI), and tested in K–5 elementary school classrooms. K–5 teachers are surveyed and then interviewed regarding sound and noise control strategies employed in their respective classrooms, by academic subject; AV media; and teacher/student configuration. Reported outcomes are compared with STI contours. The requisite background noise level and room impulse response measurements are made using a binaural acoustic mannequin at each listening position. It is expected that STI will vary among listening positions, and by academic subject and teacher/student configuration. Accordingly, scoring for the classroom will vary. Results are discussed, and based on the findings, passive means are suggested for improving STI in the classrooms studied.

3:35

1pAA10. Two challenges to achieving acoustical privacy in open offices: Going green in Switzerland. David A. Conant and William Chu (McKay Conant Hoover Inc., 5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362, DConant@MCHinc.com)

An unusual human resources office building under construction in Basel, Switzerland presented several acoustical challenges stimulated by both the architect and owner. Extensive studies demonstrated that without electronic speech sound masking, normal privacy would likely be unachievable until the receiver moved at least one floor level away from the talker. Still, the owner opted to provide no accommodation for masking for reasons we have not encountered before. Novel designs arose for absorptive chilled beams including a glass roof with chilled beam baffles. The design approaches and RASTI modeling via CATT-Acoustic are described.

3:50

1pAA11. How easy is it being green? Jesse J. Ehnert (Arpeggio Acoust. Consulting, LLC, 1947 Aspen Dr. NE, Atlanta, GA 30345, jehnert@arpeggioacoustics.com)

In 2007, the United States Green Building Council (USGBC) issued its Leadership in Energy and Environmental Design (LEED) for Schools Reference Guide defining requirements for certification which included classroom acoustical criteria. These criteria, which were based partially on the ANSI S12.60-2002 classroom acoustics standard, defined certification prerequisites (as well as additional credit requirements) in terms of three areas: reverberation time, sound isolation, and background noise level. In response to several issues related to compliance with these requirements, in 2008 the USGBC issued a Performance/Intent Equivalent Compliance Path (PIEACP) to provide alternate, and somewhat easier, means to meet LEED acoustic criteria. This presentation will weigh the potentially impractical aspects of the original guidelines and the possibly ineffective aspects of the PIEACP and will discuss the emergent areas of resistance on the part of designers and owners in relation to acoustical requirements in the context of several examples. It is the hope that properly addressing these points of resistance, and their inherent economic motivations may, perhaps, lead to a more reasonable, coherent, and enforceable set of criteria to achieve LEED certification.

4:05

1pAA12. Building ventilation and acoustics for people who don't know much about building ventilation. Michael Ermann (School of Architecture and Design, Virginia Tech, 201 Cowgill Hall, Blacksburg, VA 24061-0205, mermann@vt.edu)

The architectural composition required for building ventilation, used both for low energy cooling and improved air quality, can be anathema to acoustical goals of speech privacy and noise control. This paper presents a short tutorial on the basics of cross ventilation, stack ventilation, comfort ventilation, and indoor air quality as it relates to climate, building type, and indoor pollutants. It is geared to those without significant prior knowledge and follows a similar tutorial on geothermal systems, presented at the Miami ASA conference.

4:20—4:50 Panel Discussion

Session 1pABa

Animal Bioacoustics and Psychological and Physiological Acoustics: Fish Bioacoustics: Sensory Biology, Sound Production, and Behavior of Acoustic Communication in Fishes II

Joseph A. Sisneros, Cochair
Dept. of Psychology, Univ. of Washington, Seattle, WA 98195

Richard R. Fay, Cochair
Parmly Hearing Inst., Loyola Univ., Chicago, IL 60626

David Zeddies, Cochair
Marine Acoustics Inc., 4100 Fairfax Dr., Arlington, VA 22203

Invited Papers

1:00

1pABa1. The swimming motions of fish as hydroacoustic signals. Sheryl Coombs (Dept. of Biological Sci. , Bowling Green State Univ., Bowling Green, OH 43402)

Discussions about bioacoustic communication in fishes largely neglect the hydroacoustic fields produced by the swimming motions of fish, focusing instead on sounds produced by stridulation (the rubbing of body parts) and the drumming of swimbladder walls by surrounding muscles. This talk will describe some of the temporal and spatial characteristics of swimming-generated signals, the evidence for their ability to stimulate both the lateral line and auditory systems of fish, and their potential role in behaviorally-relevant contexts, including courtship and spawning, predator avoidance by fish schools, and spatial exploration of novel environments.

1:15

1pABa2. Auditory research from the sleeper goby to zebrafish. Zhongmin Lu (Dept. of Biology, Univ. of Miami, 1301 Memorial Dr., Coral Gables, FL 33146)

This presentation covers neural coding of acoustic particle motion by a hearing generalist (the sleeper goby, *Dominator latifons*) and research transition to a hearing specialist (the zebrafish, *Danio rerio*). Unlike sound pressure, particle motion, a component associated with sound pressure in any sound wave, provides the directional information about sound source. In recent years progress has been made in neural representation of acoustic particle motion by several hearing generalists that are primarily sensitive to the particle motion mode of sound. Here a review will be given on single-unit physiology of octavolateral afferent and efferent neurons as well as neurons in the descending octaval nucleus of the sleeper goby in response to linear acceleration that simulates underwater acoustic particle motion, providing insights into their functions. In the second half of the presentation, two preliminary projects in my lab will be introduced using the zebrafish that has become an increasingly important vertebrate model organism for biomedical research. The projects are development of the auditory function of zebrafish and alcohol effects on the octavolateral system of zebrafish. Advantages and disadvantages of different fish models will be discussed. [Work supported by NIH/NIDCD R01DC003275 and Gabelli Fellowship.]

1:30

1pABa3. Evidence of binaural directional computations in the auditory pathway of a teleost fish. Peggy Edds-Walton (Neurosci. Inst., Marine Biol. Lab, Woods Hole, MA 02543, plewalton@yahoo.com) and Richard Fay (Loyola Univ., Chicago, IL 60626)

Although binaural processing is known to be critical for sound source localization in terrestrial vertebrates, binaural processing for sound source localization in fishes has not been demonstrated. Previous physiological and neuroanatomical studies have indicated that interaural convergence is present in auditory pathway of oyster toadfish (*Opsanus tau*). In this study, we introduce a reversible method, otolith tipping, to evaluate the presence of binaural inputs. The normal directional response pattern (DRP) was recorded extracellularly for auditory cells in the first-order descending octaval nucleus (DON) or the midbrain torus semicircularis (TS) using directional stimuli in the horizontal and midsagittal planes. The DRP was re-evaluated while gently tipping the ipsilateral or contralateral otolith of the saccule to alter the normal relationship with the auditory epithelium. The DRP was also evaluated post-tipping to ensure no damage had occurred. In the DON, ipsilateral or contralateral tipping most often altered or eliminated spike activity; in the TS, tipping most often caused a change in the shape or orientation of the DRP. The data indicate complex interactions of excitatory and inhibitory inputs, which appear to be associated with the convergence of binaural inputs in the DON and TS. [Work supported by the NIH.]

1:45

1pABa4. Determining underwater auditory thresholds in fish based on particle motion. Mardi C. Hastings (Appl. Res. Lab., Penn State Univ., P. O. Box 30, State College, PA 16804, mch26@psu.edu)

Acoustic particle motion is believed to play a significant role in the auditory scene of fishes even though most all hearing sensitivity data are measured with respect to sound pressure level. Auditory motion thresholds have been measured in air for goldfish (*Carassius auratus*), oscars (*Astronotus ocellatus*), and toadfish (*Opsanus tau*) by moving the fish's body using a carefully calibrated shaker table [R. R. Fay, Science **225**, 951–954 (1984); Z. Lu, A. N. Popper and R. R. Fay, J. Comp. Physiol. A **179**, 227–233 (1996); R. R. Fay and P. L. Edds-Walton, Hearing Res. **111**, 1–21 (1997)]. These responses depend on both direction and amplitude of the motion. Thus, underwater particle motion thresholds must include both direction and amplitude, but most experimental setups are not designed for this purpose. Passive and active methods to control the direction and amplitude of acoustic particle motion underwater are presented. It is not enough just to measure the local particle motion in a small tank and report its amplitude as a threshold, especially when incidental sound pressure could be contributing to the auditory response and the orientation of the fish with respect to the direction of the particle motion vector is ill defined.

2:00

1pABa5. Peripheral neural coding strategies for spectral analysis and sound source location in the nonteleost bony fish, *Acipenser fulvescens*. Michaela Meyer, Arthur N. Popper (Dept. of Biology, Univ. of Maryland, College Park, MD 20742, michaela_meyer@meei.harvard.edu), and Richard R. Fay (Loyola Univ. of Chicago, Chicago, IL 60626)

Coding mechanisms for spectral analysis and sound source location were investigated in *Acipenser fulvescens*, the lake sturgeon. *A. fulvescens* belongs to one of the few extant nonteleost ray-finned fishes, with a phylogenetic history that dates back about 200 million years. A shaker system was used to simulate the particle motion component of sound during electrophysiological recordings of isolated single units from the eighth nerve. Data were compared to teleosts and land vertebrates. Peripheral coding strategies for spectral analysis and sound source location in *A. fulvescens* resembled those found in teleosts. Frequency characteristics generally resembled that of low-frequency afferents of other fishes and land vertebrates and the auditory periphery in *A. fulvescens* appears to be well suited to encode the intensity of sound. Eighth nerve afferents responded to directional stimuli in a cosinelike manner (as in teleosts). However, differences from teleosts were found that may have implications for the mechanisms for sound source location in azimuth. The common physiological characteristics among *A. fulvescens*, teleosts, and land vertebrates may reflect important functions of the auditory system that have been conserved throughout the evolution of vertebrates.

2:15

1pABa6. Auditory saccular potentials of the little skate, *Raja erinacea*. Joseph A. Sisneros (Dept. of Psych., Univ. of Washington, 337 Guthrie Hall, Seattle, WA 98195) and Richard R. Fay (Loyola Univ. Chicago, Chicago, IL 60626)

Previous behavioral studies have shown that elasmobranch fishes can be attracted to low-frequency sound sources (< 80 Hz) in their natural habitat, but relatively few studies have characterized the auditory response properties of the elasmobranch inner ear to low-frequency sounds. Here, we examine the response properties of evoked saccular potentials from the skate inner ear to determine the auditory threshold sensitivity and frequency response of saccular hair cells to low-frequency stimuli. Saccular potentials were recorded from the medial region of the saccule while sound was generated using a shaker table designed to mimic the particle motion component of sound. Four test frequencies (50, 64, 84, and 100 Hz) and 11 directions were used to characterize the displacement sensitivity, frequency response, and directional response properties of the skate saccule. Saccular potentials were evoked and measured at twice the stimulus frequency *in vivo* using a wave analyzer, while stimuli were generated via the shaker table system. The right and left saccule appeared to have an omnidirectional response based on our initial measurements, but the frequency response of the skate saccule had lower displacement thresholds at 84 and 100 Hz compared to stimuli at 50 and 64 Hz.

2:30

1pABa7. Sensitivity of utricular afferent fibers to intraspecific calling via inductive neural telemetry in free ranging oyster toadfish, *Opsanus tau*. Allen F. Mensinger (Biology Dept., Univ. of Minnesota Duluth, 1035 Kirby Dr. Duluth, MN 55812, amensing@d.umn.edu) and Karen P. Maruska (Stanford Univ., Stanford, CA 94305)

Male toadfish acoustically attract females to nesting sites by producing boat-whistle calls. To determine how the fish localize sound underwater, inductive neural telemetry was used to record from microwire electrodes chronically implanted into the utricular nerve of the toadfish, *Opsanus tau*. The telemetry tag allowed both laboratory and field monitoring of unrestrained, naturally behaving fish. The sensitivity of utricular afferent nerve fibers to male toadfish boat whistle calls and pure tone stimulus was determined in water depths ranging from 0.4 to 1.0 m via underwater speakers. Hydrophones were placed approximately 5 cm above the otoliths to determine the physical characteristics of the sound impinging on the utricle. Approximately 75% of the afferent fibers exhibited increased firing to sound stimulus, with many showing directional sensitivity. Thus, toadfish may be able to use information from the utricular afferents to localize sound underwater. [Work supported by NSF Grant 0316130.]

2:45

1pABa8. Phylogenetic analysis of high-frequency hearing in teleost fishes. Dennis M. Higgs (Dept. Biological Sci. , Univ. of Windsor, 401 Sunset, Windsor ON N9B 3P4, Canada, dhiggs@uwindsor.ca)

The evolutionary drivers responsible for the selection of high-frequency hearing in teleosts remain unclear. It is generally accepted that low-frequency hearing is the basal (plesiomorphic) condition, with specializations evolving to allow expansion of hearing bandwidth. While selective forces, usually habitat-based, have been proposed to explain the evolution of high-frequency hearing, phylogenetic analysis is currently lacking. The current study examines all available teleost hearing abilities in relation to habitat parameters,

salinity, and maximum depth of occurrence of each species. There was no statistical correlation between any of these parameters and maximum frequency of detection and only a weak relationship to best frequency. Phylogenetic position, at the subdivision/superorder level, did significantly predict both maximum and best frequency of detection, but there was no consistent pattern of high-frequency hearing within the Teleostei, suggesting independent evolution of this ability. These trends were also consistent at the family level. The current reanalysis of available data therefore suggests little evidence for the habitat-based hypotheses of high-frequency hearing evolution in the Teleostei. Interesting families and approaches will be highlighted in the current talk in an attempt to foster a more systematic approach to future studies of hearing in fish. [Work supported by NSERC.]

3:00—3:20 Break

3:20

1pABa9. Repetition is the branding iron of knowledge: Lessons from the mulitple origins of enhanced hearing. Christopher B. Braun (Dept. of Psych., Hunter College, 695 Park Ave., New York, NY 10021)

Enhanced hearing, the ability to detect minute fluctuations in ambient pressure, has evolved independently or nearly so within every major group of vertebrates. Within Teleost fishes, enhanced hearing has evolved perhaps as many as 20 times. Within one subset of cichlid fishes, the Malagasy and south Asian Etroplinae + Ptychochrominae, several lineages appear to have independently evolved morphological specializations that might enhance hearing. There are approximately 25 species of cichlids on Madagascar and all are narrowly endemic with well-defined habitats ranging from shallow floodplain lakes to rheophilic habitats. The gas bladders of individual species range from unspecialized (*Katria katria*) to highly modified (*Paretroplus sp*). All species of Paretroplus possess paired intracranial auditory bulla, but individual species vary in the degree of gas bladder modification (e.g., tunic stiffness, diverticula structure). This diversity presents a unique opportunity to explore the relationships among habitat noise and hydrodynamics, morphological specializations, and hearing ability within a specific phylogenetic hypothesis. This talk will present the history of hearing evolution within Malagasy cichlids, covering ecological, morphological, and functional aspects. The multiple independent origins of hearing will be emphasized as replicates of evolutionary experiments. Such replication highlights the commonalities behind the evolution of hearing in all vertebrates.

3:35

1pABa10. Comparative studies of the auditory periphery of deep-sea fish. Xiaohong Deng (Dept. of Biology and Neurosci. and Cognit. Sci. Program, Univ. of Maryland, College Park, MD 20742, xhdeng@umd.edu), H. Jochen Wagner (Univ. of Tubingen, Tubingen, D-72074, Germany), and Arthur N. Popper (Univ. of Maryland, College Park, MD 20742)

Many species of deep-sea fish have special sensory adaptations to help them survive at great depths and with lack of light. While such adaptations are best known in the visual system, it is likely that there are also auditory system adaptations that enable the fish to make use of the auditory scene. However, there are few data on the auditory system of deep-sea fish. The study compared inner ear structures between several such species. Some features in the ears in the species studied were similar to structures associated with enhanced hearing ability in the ears of some shallow water hearing specialists. These findings support the hypothesis that some deep-sea fishes have evolved specializations for hearing. We also report some features in the ears of deep-sea fishes that have never been seen in the ears of other vertebrates. These include specializations in the orientation of sensory hair cells, diversity in hair cell ciliary bundle shapes, and a number of unique otolith characteristics. While it is not possible to test hearing in deep-sea fishes, the various adaptations found suggest that at least some have evolved highly specialized adaptations to enable them to use sound in the deep sea.

3:50

1pABa11. Ontogeny of auditory saccular tuning in the plainfin midshipman fish, *Porichthys notatus*. Peter W. Alderks and Joseph A. Sisneros (Dept. of Psych., Univ. of Washington, Box 351525, Seattle, WA 98195, pwa2@u.washington.edu)

Sensory systems are important throughout an animal's life history, allowing it to detect and respond to biologically relevant stimuli important for survival and reproduction. The auditory system of the plainfin midshipman fish (*Porichthys notatus*) plays an important sensory receiving role for encoding vocal-acoustic communication signals produced by adults, but the response properties and function of this receiver system during ontogeny are less understood. Here, we examine the response properties of evoked saccular potentials in two size classes of juveniles [small = 1.9–3.0 cm standard length (SL) and large = 5.0–8.0 cm SL] and in nonreproductive adults ($>$ 9.0 cm SL) to determine the auditory sensitivity of saccular hair cells during ontogeny. Saccular potentials, maximally evoked at twice the stimulus frequency, were recorded in vivo using a lock-in amplifier while playing pure tone stimuli via an underwater speaker. Results indicate an ontogenetic retention of saccular tuning sensitivity with size/age with peak sensitivity ranging from 75 Hz (lowest frequency tested) to 185 Hz. In addition, maximum detectable frequency of the saccule also increased with size/age. These results suggest that the midshipman saccule is best suited to detect low frequency components ($<$ 105 Hz) of conspecific vocalizations throughout ontogeny.

4:05

1pABa12. Central auditory processing and neuropeptide modulation of hearing in the damselfish. Karen P. Maruska (Biology Dept., Stanford Univ., 371 Serra Mall, Stanford, CA 94305, maruska@stanford.edu) and Timothy C. Tricas (Univ. of Hawaii, Honolulu, HI 96822)

Acoustic communication is important for social behaviors in many fishes, but how the central auditory system encodes complex sounds and how neuromodulators influence hearing is poorly understood. The Hawaiian sergeant damselfish produces pulsed low-frequency low-intensity sounds during agonistic and reproductive behaviors, and provides an ideal model to study auditory processing and hormone effects in a species that does not possess special adaptations to enhance hearing. Single-cell recordings were used to (1) compare response properties and sensitivity of auditory neurons between the hindbrain and midbrain to both simple tones and playbacks

of complex conspecific sounds, and (2) test the effects of the neuromodulator gonadotropin-releasing hormone (GnRH) on midbrain auditory response properties. Auditory neurons in both the hindbrain and midbrain showed similar high phase-locking ability and best frequencies of 80–400 Hz, but midbrain neurons were more sensitive to conspecific sounds compared to hindbrain neurons. GnRH application to the midbrain showed primarily inhibitory effects on evoked auditory responses, evidenced by a GnRH-induced decrease in spike rate compared to saline controls. These data highlight the importance of using natural stimuli to test hearing abilities, and indicate that GnRH has the potential to influence central auditory processing during natural social behaviors in the damselfish.

4:20

1pABa13. Substrates for vocal-acoustic plasticity: Neuroestrogen and steroid hormone receptor expression in midshipman fish. Paul Forlano (Dept. of Neurobiology and Physio., Northwestern Univ., Evanston, IL 60208, p-forlano@northwestern.edu), Margaret Marchaterre, David Deitcher, and Andrew Bass (Cornell Univ., Ithaca, NY 14853)

In a teleost fish, the midshipman *Porichthys notatus*, androgens and estrogen rapidly modulate vocal motor patterning and contribute to seasonal changes in the frequency encoding of the eighth nerve. To identify potential sites of action of these steroids, as well as local estrogen synthesis (via conversion from testosterone by the enzyme aromatase), in situ hybridization was employed to identify estrogen receptor alpha (ER), androgen receptor (AR), and aromatase mRNA distribution in vocal and auditory circuitry. AR expression was more widespread than ER and found at all levels of the vocal motor pathway and in vocal-acoustic integration centers, while aromatase was particularly abundant in the vocal motor system. Both AR and ER were identified in midbrain and forebrain auditory nuclei and in the saccular epithelium of the inner ear, the main auditory end organ in midshipman and most teleosts. Aromatase was found in ganglion nerve cells of the eighth nerve, suggesting that the ear itself is a local source of estrogen driving peripheral auditory plasticity. In sum, expression of steroid receptors in vocal and auditory nuclei, and the inner ear, support neurophysiological evidence for the role of estrogen and androgens as modulators of vocal output and auditory encoding in midshipman fish. [Work supported by NIMH, NIH, and NSF.]

4:35

1pABa14. Otolith crystallization in salmonid fishes: Possible causes and consequences for the fish. Allison B. Coffin (Dept. of Otolaryngol.-H&S Surgery, Univ. of Washington, Box 357923, Seattle, WA 98195, coffina@u.washington.edu)

The primary hearing organ in most fishes is the saccule, which contains a calcareous otolith composed of aragonite and its associated sensory epithelium. In some fish populations, particularly salmonids, the aragonite in the saccular otolith is partially or mostly replaced by vaterite, a phenomenon known as otolith crystallization. Despite the relatively common occurrence of otolith crystallization, the cause of this switch in calcium carbonate type and the consequences for the fish are not understood. The present study looks at both proximal and distal mechanisms that may contribute to otolith crystallization in various salmonid species, particularly with regard to environmental stress and physiological consequences of that stress. Structural and functional data on the ears of salmonids with crystallized otoliths will also be presented. Auditory evoked potential recordings from juvenile chinook salmon demonstrated that fish with at least one vaterite otolith have significantly increased hearing thresholds compared to conspecifics with two aragonite otoliths. These findings suggest that otolith crystallization can have potential negative consequences for the animal.

MONDAY AFTERNOON, 18 MAY 2009 COUNCIL SUITE, 2:30 TO 4:45 P.M.

Session 1pABb

Animal Bioacoustics: An Integration of Bioacoustics, Neuronal Responses, and Behavior II

Terry Takahashi, Chair
Inst. of Neuroscience, Univ. of Oregon, Eugene, OR 97403

Chair's Introduction—2:30

Contributed Papers

2:35

1pABb1. Ultrasonic upward FM vocalization of Mongolian gerbils, *Meriones unguiculatus*. Kohei Nishiyama, Kohta Kobayasi, and Hiroshi Riquimaroux (Faculty of Life and Medical Sci. , Doshisha Univ., 1-3 Miyakotani Tatara, Kyotanabe 610-0321, Japan, hrikimar@mail.doshisha.ac.jp)

Mongolian gerbils, *Meriones unguiculatus*, have been used for auditory physiological studies. However, few studies have investigated precise relationships between a specific vocalization and a specific behavior. The purpose of this study was to examine the process to produce a specific vocalization: Greeting call. Their vocalizations and behaviors were separately recorded when two gerbils are put in the same experimental chamber. The sound data were synchronized with movie data to analyze their movement loci by using behavioral analysis software. The results showed that gerbils tended to emit ultrasonic upward FM calls, Greeting calls (average frequency range: 29.4 to 34.9 kHz, duration: 27.0 ms, $n=$ 1185 calls), when the distance between two gerbils was close. About 45% of the calls were emitted when the distance was less than 10 cm. Moreover, in the situation where one gerbil quickly approached to the other, timings of Greeting call emission and footsteps of the approaching gerbil appeared to be synchronized. However, they did not vocalize at all when they were isolated, even when they moved in the same way. We may conclude that the repetitive upward FM ultrasonic emissions were not by-products of their movements but for their communication. [Research supported by ONR grant.]

2:50

1pABb2. Jamming of bat sonar by meadow katydids (*Conocephalus*). James A Simmons (Dept. of Neurosci., Box G-LN, Brown Univ., Providence, RI 02912)

Meadow katydids (Orthoptera: family Tettigoniidae, genus *Conocephalus*) are small (1–2 cm) grasshopperlike insects that congregate in late summer on low bushes and tall grass near ponds where males then emit calls to attract females. These sounds are entirely ultrasonic, consisting of sequences of brief (20–25 ms) bursts, repeated at 30 ms intervals, of 10–20 clicks separated by 1–1.5 ms within each burst. Each click covers frequencies of 20–90 kHz. The same frequencies are used for sonar by big brown bats (*Eptesicus fuscus*), which frequent the same ponds while hunting for prey and nearly always fly in audible range of the meadow katydids but do not attempt to capture any of them. Calls of several male insects usually overlap at slightly different inter-click intervals so that the acoustic waveform wavers, thus creating moving virtual acoustic sources for the bat's binaural listening. Aggregate calling is virtually continuous at sound pressures exceeding 90–100 dB SPL at distances of 1–2 m. Neuronal recovery-times in the bat's auditory system are ~0.5 ms to accommodate reception of closely-spaced echolocation emissions and multiple echoes, so click bursts will reliably evoke volleys of responses at successive stages of the bat's auditory pathway, preempting biosonar echoes received by approaching bats.

3:05

1pABb3. FM echolocating bats avoid pulse-echo ambiguity by shifting broadcast frequencies. Shizuko Hiryu (Doshisha Univ., 1-3 Tatara Miyakotani, Kyotanabe, Kyoto 610-0321, Japan), Mary Bates, James A. Simmons (Brown Univ., Providence, RI 02912), and Hiroshi Riquimaroux (Doshisha Univ., Kyotanabe, Kyoto 610-0321, Japan)

Big brown bats pursue flying insects in open areas, but they also hunt for prey near vegetation. The proximity of obstacles requires emission of sounds at short interpulse intervals (IPIs) to guide collision-avoidance maneuvers, yet the bat has to locate obstacles distributed in depth for path planning, which requires longer IPIs. If IPIs are too short, the stream of echoes for one sound will still be arriving when the second sound is emitted, creating ambiguity about matching echoes to broadcasts. We investigated how bats cope with pulse-echo ambiguity by recording sounds with a miniature radio microphone (Telemike) carried during flights through a dense array of obstacles (vertically hanging chains) that were distributed in depth. Array density forced bats to emit FM sounds at short intervals (strobe groups) with consequent pulse-echo ambiguity. We found that when ambiguity occurred, bats shifted the FM sounds in strobe groups upward and downward in frequency by several kiloHertz to distinguish each broadcast's echoes. Overlap of echo streams from the bat's own FM broadcasts causes extreme clutter interference, and bats react by slightly shifting FM frequencies to distinguish one sound's echo stream from another sound's. [Research supported by an ONR grant.]

3:20

1pABb4. Relationships between wingbeat and pulse-emission timings in three echolocating bats, *Rhinolophus ferrumequinum, Pipistrellus abramus*, and *Miniopterus fuliginosus*. Yu Watanabe, Shizuko Hiryu, Hiroshi Riquimaroux, and Yoshiaki Watanabe (Doshisha Univ., 1-3 Tatara Miyakotani, Kyotanabe, Kyoto 610-0321, Japan)

To investigate relationships between flight and echolocation behaviors in one CF-FM (*Rhinolophus ferrumequinum*) and two FM (*Pipistrellus abramus* and *Miniopterus fuliginosus*) Japanese echolocating bats, relationships between wingbeat and pulse-emission timings during landing flight were examined. Wingbeat timings were measured by a dual high speed video camera system (500 fps) and pulse-emission timings were precisely recorded by a wireless telemetry microphone mounted on the bats' head. The *R. ferrumequinum* and *P. abramus* emitted multiple pulses in a burst (pulse trains) per single wingbeat cycle, and pulse-emission timings were perfectly synchronized with their wingbeat, indicating that pulse-emission timings were affected by wingbeat. On the other hand, relationships between wingbeat and pulse-emission timings in *M. fuliginosus* were different compared with two other bats. The *M. fuliginosus* never emitted pulse trains, but proportionally increased emission rate of single pulse after approach phase started. As a consequent, pulse-emission timings were controlled regardless of wingbeat, and synchronization between wingbeat and pulse-emission timings was occasionally disturbed during approach phase. These results suggest that we need further discussion about the reason why synchronization between flight and echolocation behaviors appears to be formed by species-specific way among these bats. [Work supported by MEXT of Japan.]

3:35

1pABb5. Multisensory integration in the tadpole medulla arises from discrete and overlapping regions of endorgan projection. Andrea M. Simmons, Victoria Flores, and Seth S. Horowitz (Dept. of Psych., Brown Univ., 89 Waterman St., Providence, RI 02912)

Vestibular and auditory nuclei in the medulla are inherently multimodal in developing bullfrogs (Rana catesbeiana), responding to pressure, particle motion, and lateral line input. While brainstem and midbrain connectivity of these nuclei have been well mapped across metamorphic development, developmental changes in projection patterns from individual endorgans in the otic capsule (sacculus, utricle, lagena, and amphibian and basilar papilla) to the medulla are less well understood. Using small (26 nl) focal injections of neurobiotin in preganglionic branchlets for specific endorgans, we were able to map input from the separate peripheral organs to different medullary nuclei (medial, superior and lateral vestibular nuclei, the dorsal medullary nucleus, and the anterior lateral line nucleus). Some secondary transsynaptic label of the reticular gray and cerebellum was also noted. Projection patterns are correlated with different patterns of medullary responses to vibratory and to pressure stimulation. Endorgan projection patterns appear relatively stable across metamorphic development. These data suggest that changes in functional response properties in the medulla across development do not reflect rewiring of inner ear input, but instead depend on changes in connectivity between the medullary nuclei and their more central target zones. [Work funded by NIDCD and the Rhode Island Space Grant Consortium.]

3:50—4:00 Break

4:00—4:45 Panel Discussion

Session 1pAO

Acoustical Oceanography, Underwater Acoustics and Signal Processing in Acoustics: Temporal and Spatial Field Coherence Applied to Ocean Sensing: Measurement, Theory, and Modeling II

Kyle M. Becker, Cochair
Applied Research Labs., Pennsylvania state Univ., State College, PA 16804-0030

Michael G. Brown, Cochair
RSMAS-AMP, Univ. of Miami, Miami, FL 33149-1098

Chair's Introduction—1:00

Invited Papers

1:05

1pAO1. Spatial and temporal coherence of scattered field from large fish groups instantaneously imaged with ocean acoustics waveguide remote sensing (OAWRS). Purnima Ratilal, Mark Andrews, Zheng Gong (Dept of Elec. and Comp. Eng., Northeastern Univ., 360 Huntington Ave., Boston, MA 02115, purnima@ece.neu.edu), and Nicholas Makris (MIT, Cambridge, MA 02139)

Wide area sonar system, such as OAWRS (ocean acoustics waveguide remote sensing), can instantaneously image large shoals of fish spanning tens of kilometers from long ranges in an ocean waveguide. An analytic and numerical model is developed for the spatial and temporal coherence of the scattered field from a large group of fish in a shallow water waveguide. The temporal coherence of the received signal depends on both the temporal coherence of the random waveguide through the medium Green's function and the correlation time scale for fish group distribution. The spatial coherence of the received signal depends on both the spatial correlation scale of the random waveguide inhomogeneities and that for the imaged fish group distribution. The temporal and spatial correlation scale for the fish group distribution are determined from experimental data acquired in the Gulf of Maine during the 2006 Ocean Acoustics Waveguide Remote Sensing (OAWRS) Experiment. The effect of temporal and spatial coherence of the scattered field on subsequent coherent array processing and matched-filtering is discussed.

1:25

1pAO2. Investigation of the impact of water column variability on acoustic inversions in a shallow water ocean. Ying-Tsong Lin, James F. Lynch, and Timothy F. Duda (Woods Hole Oceanograph. Inst., AOPE Dept. MS 11, Woods Hole, MA 02543)

Underwater sound propagation in a rapidly evolving shallow-water ocean has high temporal and spatial variability. To ensure the performance of acoustic inversion techniques to locate a sound source or estimate the sea-bottom parameters, it is necessary to take into account the water column variability, which can be done by introducing the empirical orthogonal functions (EOFs), which describe the water column variability, into the environmental model used for the inversions. However, in a range-dependent and time varying ocean, one needs to regularly update the EOFs according to the scales of environmental, spatial, and temporal coherence, or the EOFs may not be the most efficient and optimal vector bases to decompose the dynamic ocean field. In this paper, the environmental data collected during the New Jersey Shallow Water 2006 (SW06) experiment are analyzed to check the environmental coherence scales in the experimental site. The analysis results provide a guideline as to when and where one should update the EOF bases. The observation of the SW06 acoustic signal coherence is reported and correlated to the environmental variability. Finally, acoustic inversions using the SW06 data are performed, and the effects of using insufficient water-column EOFs are investigated.

1:45

1pAO3. Extracting waveguide parameters from the coherent field measured on a synthetic aperture horizontal array. Kyle M Becker (Pennsylvania State Univ./Appl. Res. Lab., P.O. Box 30, State College, PA 16804-0030)

For a time-harmonic source in a stratified waveguide, the acoustic field can be approximated by a coherent sum of modes propagating in range. In range-dependent waveguides, the phase contribution from mode n is the product of a range-dependent wave number $k_n(r)$ and range. At any given range, k_n is dependent on water depth and the depth-dependent geoacoustic parameters of both the water column and the seabed. To exploit these relationships, inversion algorithms have been developed for estimating range-dependent seabed properties from the acoustic field measured on synthetic aperture horizontal arrays. These methods are based on inferring information about the waveguide environment by extracting information contained in the phase of the acoustic signal as a function of range. In this talk we examine the individual contributions to the total phase from each of the propagating modes based on changes in depth, sediment sound speed, and water column properties. Sensitivity of individual modes to each of these changes will be examined along with their contribution to the total field. Issues associated with estimating modal content from field data and their use in inverse problems will then be discussed.

2:05

1pAO4. Passive remote sensing through cross-correlation of nondiffuse ambient noise. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA/Earth System Res. Lab., Boulder, CO 80305, Oleg.Godin@noaa.gov)

Under rather restrictive assumptions about ambient noise field and its sources, it has been demonstrated theoretically by a number of authors that one can retrieve the exact acoustic Green's function (GF) of an inhomogeneous medium from the noise cross-correlation function (NCCF). These assumptions are rarely valid for ambient noise in the ocean. In this paper, asymptotic approaches are used to investigate the information content of two-point correlation functions of ambient noise and their possible use for passive remote sensing of the environment under more general conditions, when there exists no simple, exact, local relation between NCCF and deterministic GFs. For noise sources distributed on a surface or in a volume of a moving or motionless, inhomogeneous fluid, a relation is established between contributions of ray arrivals to NCCF and GFs. A similar relation is derived for contributions to NCCF and GFs due to adiabatic normal modes in a horizontally inhomogeneous waveguide. The impact of nonuniformity of the spatial distribution of noise sources on accuracy of passive measurements of acoustic travel times is quantified. Effects of finite correlation length of noise sources on GF retrieval from ambient noise are discussed. [Work supported by ONR.]

2:20

1pAO5. Extension of path-integral theory to shallow-water coherence. Peter C. Mignerey (Acoust. Div. 7120, Naval Res. Lab., Washington, DC 20375, peter.mignerey@nrl.navy.mil)

A path-integral approach to the mutual coherence function of acoustic signals propagating in deep water was developed in the late 1970's. This theory requires a single dominant ray path that interacts weakly with the boundaries, a restriction which has precluded direct application to acoustic propagation in shallow water. Here, path-integral theory is extended to the mutual coherence of mode amplitudes defined over the horizontal plane. In this domain, the mode amplitudes follow a principal ray that has no interaction with lateral boundaries. The mutual coherence is found to be related to the spectrum of sound-speed fluctuations via a Fourier transform along a direction transverse to the acoustic propagation path. For sound-speed fluctuations generated by a linear internal-wave field, solutions are obtained that predict the mutual coherence with no free parameters. Such solutions were computed for a winter environment existent during an experiment that took place on the New Jersey shelf in December 2003. Comparison between theoretical and measured coherence is encouraging with respect to both the power-law behavior at short separations, and the transition to saturation at large separations. [Work supported by the Office of Naval Research.]

2:35

1pAO6. Effect of the internal tide on midfrequency transmission loss in the Shallow Water 2006 Experiment. Jie Yang, Daniel Rouseff, Dajun Tang, and Frank S. Henyey (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

Nonlinear internal waves depress the high-gradient region of the thermocline. After the waves have passed, it may take several hours for the thermocline to rise to its prewave level. To examine consequent acoustic effects, two acoustic data sets collected 18–19 August in SW06 were analyzed with one at a fixed range of 550 m and the other along a tow track out to 8.1 km. Both data sets allow the effects of the rising thermocline on acoustic propagation to be examined over a range of temporal and spatial scales. Signals transmitted are in the midfrequency band 1.5–10.5 kHz. The gradual rising of the thermocline is shown to change acoustic intensity by 5 and 2 dB for the fixed and towed source data, respectively. Using nearby oceanographic moorings, a simple plane wave ocean model is developed to provide time-dependent environmental input to broadband acoustic modeling assuming range-independency. Modeling results are shown to produce mean transmission loss results consistent with experimental observations at both the short and the long ranges. The results suggest that the rising thermocline effect of nonlinear internal waves on mean transmission loss is observable, significant, and predictable. The result has implications on geoacoustic inversion. [Work supported by ONR.]

2:50

1pAO7. Inversions from source of opportunity data collected on horizontal and vertical line arrays. Steven A. Stotts, Robert A. Koch, Sumedh M. Joshi, Vian T. Nguyen, and David P. Knobles (Appl. Res. Labs., The Univ. of Texas at Austin, 10000 Burnet Rd., Austin, TX 78758, stotts @arlut.utexas.edu)

Environmental and source track parameter values were extracted by coherent spatial and temporal field processing of data from a ship of opportunity source that were recorded on collocated horizontal and vertical arrays. The data were collected during the Shallow Water 2006 (SW06) experiment in a continental shelf, sandy sediment environment with a water depth of approximately 70 m. Inversion solutions for environmental and source track parameter values were obtained from both conventionally and adaptively beamformed subaperture data over a 5-km range scale. The inversion solution for the data from the source to the closest point of approach (CPA) is presented and compared to the prior information on the seabed parameter values provided by geophysical measurements and inversions using other measured acoustic data. The coherent received field structure for the source departing from CPA is compared to the model predictions obtained from the inversion solution for the seabed parameter values. Inversion solutions from the vertical line array data are compared to those obtained from the horizontal line array data. [This work is supported in part by ONR.]

3:05—3:30 Break

3:30

1pAO8. The evolution of vertical spatial coherence with range from source. Peter H. Dahl (Appl. Phys. Lab. and Dept. of Mech. Eng., Univ. of Washington, Seattle, WA 98105), Dajun Tang (Univ. of Washington, Seattle, WA 98105), and Jee Woong Choi (Hanyang Univ., Ansan, Korea)

Vertical spatial coherence for shallow water propagation at frequencies 1–10 kHz is studied as function of range (50 to 5000 m), as part of the Shallow-Water 2006 program that took place off the coast of New Jersey in August 2006 in waters 80 m deep. An acoustic source was deployed from the R/V Knorr at depths 30 and 40 m and signals were recorded on a moored receiving system consisting of two 1.4 m long vertical line arrays (VLA) centered at depths 25 and 50 m. At all ranges, spatial coherence, Γ (normalized spatial correlation), is locally stationary and depends on element vertical separation d up to the maximum kd (59) afforded by the VLA, where k is acoustic wave number. For range normalized by depth, r^*, less than about 10, Γ is oscillatory, with non-zero imaginary part, reflecting the inclusion of multipaths for which no single path dominates. For r^* greater than 10, Γ tends to exhibit a monotonic decay with kd and the imaginary part vanishes reflecting symmetry about 0 deg vertical arrival angle. The coherence also increases with r^* reflecting the change in modal structure. [Research supported by ONR Ocean Acoustics.]

3:45

1pAO9. Exploiting correlation in passive sonar Bayesian localization techniques. Colin W. Jemmott, R. Lee Culver (Penn State Graduate Program in Acoust., P.O. Box 30, State College, PA 16804), N. K. Bose, and B. E. Bissinger (Penn State, State College, PA 16804)

Low frequency acoustic signals in shallow water are strongly affected by interference between multiple paths resulting from boundary interactions. As the acoustic source moves through this interference pattern, the spatial variation in transmission loss can result in strong temporal modulation of the received signal, which can be used to localize the source. Acoustic propagation models can produce accurate transmission loss field predictions, but are sensitive to ocean environmental parameters such as bottom composition, bathymetry, and sound speed profile. If the uncertainty in the undersea environment can be described by probability density functions of these param-

eters, Monte Carlo forward models can be used to produce an ensemble of possible transmission loss realizations. A probabilistic model representing this ensemble must include a density function of transmission loss at each location, as well as correlation of transmission loss between locations. In addition, the choice of probabilistic model has a large impact on the form of the resulting Bayesian localization algorithm. Previous results have shown that including spatial correlation of transmission loss can result in improved localization. This talk will introduce a non-Gaussian probabilistic model for representing uncertainty in transmission loss predictions that includes correlation, and discuss the resulting recursive Bayesian localization algorithm.

4:00

1pAO10. Effect of noise sources motion on the performance of passive imaging from ambient noise cross-correlation. Karim G. Sabra (School of Mech. Eng., Georgia Inst. of Technol., 771 Ferst Dr. NW, Atlanta, GA 30332-0405)

It has been demonstrated theoretically and experimentally that an estimate of the Green's function between two receivers can be obtained from the long-time average of the cross-correlation of ambient noise at these two receivers in various environments and frequency ranges of interest: ultrasonics, underwater acoustics, and seismology. The coherent wavefronts emerge from a correlation process that accumulates contributions over time from noise sources whose propagation path passes through both receivers. Previous theoretical studies of the performance of this noise-based passive imaging technique assumed that the noise sources remain stationary. The influence of the motion of the noise sources (Doppler effect) on this passive imaging technique was investigated theoretically in both free spaces using a stationary phase approximation and extended to waveguide environments. Although the Doppler effect typically degrades the performance of standard wideband coherent processing schemes, such as time-delay beamforming of fast moving sources, it was found that the Green's function estimates extracted from ambient noise cross-correlations are not affected by the Doppler effect. Practical implications for at-sea passive tomography experiments using shipping noise will be discussed. [Work supported by ONR, Code 321.]

4:15

1pAO11. Measurements of acoustic propagation with accompanying high-resolution sound speed structure containing internal waves. Frank S. Henyey, Kevin L. Williams, and Dajun Tang (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, frank@apl.washington.edu)

During the Shallow Water 2006 experiment, simultaneous measurements were made of the sound speed structure associated with nonlinear internal waves and acoustic propagation at frequencies of 2–10 kHz over a 1 km path. The internal waves were measured by a towed CTD chain in order to get high resolution. These measurements were coordinated so that the nonlinear waves can be interpolated onto the acoustic path, allowing predictions of their effects on the acoustics. An internal wave train was measured that passed the acoustic path on August 13. When the wave train was in between the sound source and receiver, distinctive arrival time oscillations on three acoustic paths were measured, which are all rays having an upper turning point. Using the CTD chain data, a deterministic explanation is given to the arrival time oscillations.

MONDAY AFTERNOON, 18 MAY 2009 PAVILION EAST, 1:00 TO 4:50 P.M.

Session 1pBB

Biomedical Ultrasound/Bioresponse to Vibration: Biological Medium Characterization and Imaging

Azzdine Y. Ammi, Chair
Oregon Health and Science Univ., Portland, OR 97239

Chair's Introduction—1:00

Contributed Papers

1:05

1pBB1. Estimation of average acoustic concentration of Chinese hamster ovary cells using ultrasonic backscatter. Jun Song and William O'Brien, Jr. (URH326, Univ. of Illinois at Urbana-Champaign, 1111 W. Nevada St., Urbana, IL 61801, junsong2@illinois.edu)

Average acoustic concentration (AAC) is one important parameter of quantitative ultrasonic assessment of soft tissues. The objective of this study is to examine the AAC of Chinese hamster ovary (CHO) cells at varied cell densities using ultrasonic backscatter. The CHO cells were evenly distributed in 2% agar and the densities ranged from 500,000 to 54,000,000 cells /mL. Measurements of backscattered ultrasound were obtained over the frequency range of 17 to 80 MHz. Estimates of Average Scatterer Diameters and AAC were made by comparing the backscattered power spectra with three scattering models: Fluid-filled Sphere model (FFSM), Spherical Gaussian model (SGM), and Concentric Sphere Model (CSM). Over the frequency range of 28 to 50 MHz, the estimated ASDs were 20.0 ± 0 μm with FFSM, 22.2 ± 6.0 μm with SGM, and 20.1 ± 6.1 μm with CSM. In the cell density range of 500,000 to 10,000,000 cells/mL, the estimated AACs increased almost linearly from 10 to 34 dB/mm^3 using FFSM, from 16 to 39 dB/mm^3 using SGM, and from 8 to 31 dB/mm^3 using CSM. [Work supported by NIH Grant CA111289.]

1:20

1pBB2. Mammary tumor classification using envelope statistics from ultrasound backscatter and the homodyned K distribution. David Hruska, William D. O'Brien, Jr., and Michael Oelze (Dept. of Elec. and Comp. Eng., Univ. of Illinois at Urbana–Champaign, 405 N. Mathews Ave., Urbana, IL 61801)

Envelope statistics from ultrasound backscatter based on the homodyned K distribution were used to characterize organizational patterns of tissue microstructure in three rodent models of breast cancer: a mouse mammary carcinoma, a mouse mammary sarcoma, and a rat fibroadenoma. The homodyned K distribution was used to model the amplitude of the envelope of backscattered ultrasound yielding two parameters to characterize tissues: one related to the effective scatterer density and the other related to periodicity of scatterer locations. Ten tumors of each of the three kinds of tumors were scanned. Envelope statistics model parameters were estimated for regions of interest measuring 1×1 mm in each of the tumor images. These estimates were then averaged together for each tumor. A linear combination of the two estimated parameters was derived for each tumor in order to obtain a lumped parameter for discriminating between the different kinds of tumors. This linear combination yielded the following (mean$\pm$SD): carcinomas 9.58 ± 1.17, sarcomas 12.01 ± 0.78, and fibroadenomas 11.66 ± 1.15. These estimates were statistically significant in distinguishing carcinomas

from sarcomas ($p<0.05$) and carcinomas from fibroadenomas ($p<0.05$), but a statistically significant difference was not observed between the sarcomas and fibroadenomas ($p=0.44$). [Work supported by NIH CA111289.]

1:35

1pBB3. On the use of coded excitation and pulse compression to reduce estimate error of average scatterer diameters obtained from ultrasonic backscatter. Jose R. Sanchez, Darren Pocci, and Michael L. Oelze (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Matthews, Urbana, IL 61801)

A technique that extends the tradeoff of variance of average scatterer diameter (ASD) estimates and the spatial resolution in quantitative ultrasound (QUS) images is proposed. Resolution enhancement compression (REC), a coded excitation and pulse compression technique, was used to enhance the −6-dB bandwidth of an ultrasonic imaging system. The objective of this study was to combine REC with QUS (REC-QUS), and evaluate and compare improvements in ASD estimates obtained using the REC technique to conventional pulsing (CP) methods. Simulations and experimental measurements were conducted with a 10 MHz weakly focused (f/4) single-element transducer (−6-dB bandwidth of 80%). Using REC, the −6-dB bandwidth was enhanced to 152%. In simulations, a homogeneous phantom with scatterers 50 μm in diam was evaluated. In experimental measurements, a tissue-mimicking phantom that contained glass spheres with scatterer diam that range from 45 to 53 μm was evaluated. Estimates of ASD were obtained by comparing the normalized backscattered power spectra to theory. Simulation results indicate that for a constant variance, a reduction in region-of-interest size from 15 CP pulse lengths to 1.5 CP pulse lengths was achieved by using REC-QUS. In experimental measurements, the estimate variance decreased by 88% by using REC-QUS over conventional QUS. [Work supported by NIH-R21-EB006741.]

1:50

1pBB4. Improvements in compression of coded excitation echoes by using a spatially varying Wiener filter. Jose R. Sanchez and Michael L. Oelze (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Matthews, Urbana, IL 61801)

Coded excitation schemes require compression of echoes to restore pulse resolution after transmission of an elongated signal. Compression schemes are often performed using a Wiener filter, which contains a smoothing parameter that controls the trade off of gain in echo signal-to-noise ratio (eSNR), sidelobe levels, and axial resolution. A Wiener filter that takes into consideration the spatially varying nature of the pulse-echo impulse response of the ultrasound imaging system within the depth of field was explored. This spatially varying Wiener filter (SVWF) generates numerous compressed outputs within the depth of field that are compounded into one final output, i.e., one final image. The final image using the SVWF has improved sidelobe levels and eSNR when compared to not using the SVWF for compression. Experimental measurements were conducted with a 2.25 MHz weakly focused (f/3) single-element transducer (−6-dB bandwidth of 75%). The imaging targets evaluated consist of planar reflectors, wire targets in water, and wire targets in urethane rubber (ATS 539 multipurpose tissue-mimicking phantom). To evaluate the performance of the SVWF, the spatial resolution, eSNR, and sidelobe-to-mainlobe ratio were evaluated from the images. Preliminary results suggest that compressing at each spatial location within the depth of field improved sidelobe levels leading to improved imaging performance. [Work supported by NIH-R21-EB006741.]

2:05

1pBB5. Extending the trade-off between spatial resolution and variance in quantitative ultrasonic backscattering imaging (QUS) using full angular spatial compounding. Roberto J. Lavarello, Jose R. Sanchez, and Michael L. Oelze (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana–Champaign, 405 North Mathews Ave., Urbana, IL 61801)

Tissue microstructure properties can be estimated using backscattered power measurements and quantitative ultrasound (QUS) techniques. The objective of this study was to evaluate the use of full angular (i.e., 360°) spatial compounding as a means of extending the trade-off between QUS variance and spatial resolution. Simulations and experimental results were conducted with a 10 MHz, *f*/4 transducer. In simulations, a synthetic phantom consisting of two eccentric cylindrical regions with Gaussian inclusions with 50 and 25 μm average scatterer diameter (ASD) values was analyzed. The use of multiple view data reduced the corresponding ASD standard deviations from 13.7 and 19.6 μm to 2.5 and 3.7 μm, respectively. In experimental measurements, a phantom that contained glass spheres with diameters between 45 and 53 μm was evaluated. The ASD standard deviation obtained using single view data was 10.4 μm. When using 32 angles of view and reducing both region-of-interest (ROI) dimensions by a factor of 2, the ASD standard deviation was reduced to 4.8 μm. The results presented here demonstrate that both QUS spatial resolution and precision can be improved simultaneously by using full angular spatial compounding. This method finds direct applicability for breast tissue characterization, for which full angular coverage is available.

2:20

1pBB6. A study of two techniques for three-dimensional ultrasonic tomography using inverse scattering. Roberto J. Lavarello and Michael L. Oelze (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana–Champaign, 405 North Mathews Ave., Urbana, IL 61801)

A majority of inverse scattering works consider 2D imaging problems. Therefore, the study of 3D inverse scattering methods is necessary for future applications of ultrasonic tomography. Two approaches were examined through simulations for image reconstruction of spherical targets. In the first approach, transducers focused on elevation were used to produce a series of 2D image slices of the object using the distorted Born iterative method (DBIM). The resulting 3D image reconstruction was rendered by stacking serial 2D slices. Typical mean average error (MAE) values in sound speed contrast reconstructions were above 25%. Reconstruction artifacts due to large multiple scattering were identified in addition to the limited spatial resolution in the vertical direction predicted by linear imaging theory. The magnitude of the artifacts decreased with decreasing focal number and also depended on properties of the imaging target (i.e., sound speed contrast and radius). In the second approach, pointlike transducers were used to produce a fully 3D DBIM image reconstruction by inverting the 3D wave equation. Reconstructions using 3D DBIM resulted in MAE values around 13%, and no noticeable artifacts when transducers were properly distributed. The dependence of reconstruction quality on transducer arrangement was studied by analyzing three different array configurations.

2:35—3:05 Break

3:05

1pBB7. Fractal ladder models and power-law wave equations. James F. Kelly (Dept. of Appl. Math., Spanagel Hall, Rm. 250, 833 Dyer Rd., Monterey, CA 93943-5216) and Robert J McGough (Michigan State Univ., East Lansing, MI 48823)

Extensive measurement of the ultrasonic attenuation coefficient in human and mammalian tissue has revealed a power-law dependence on frequency. To describe this power-law behavior in 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, the fractional partial differential equations previously proposed by [M. Caputo "Linear models of dissipation whose *Q* is almost frequency independent--II," Geophys. J. R. Astron. Soc. **13**, 529–539 (1967)] and [MG. Wismer "Finite element analysis of broadband acoustic pulses through inhomogeneous media with power-law attenuation," J. Acoust. Soc. Am. **120**, 3493–3502 (2006)] are derived. The resulting attenuation coefficient is a power-law with exponents ranging between 1 and 2, while the phase velocity is in agreement with the Kramers–Kronig relations. The power-law exponent is interpreted in terms of the mechanical structure and related to the spectral dimension of the underlying fractal geometry.

3:20

1pBB8. Shear wave elastography of cetacean tissues. Michael Gray, James Martin, and Peter Rogers (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

An ultrasound-based elastography system is being developed to noninvasively measure the complex shear speed in cetacean head tissues, including brain, jaw fat, and melon. The system instrumentation features a pair of dual-element confocal ultrasound transducers, one of which is used to remotely generate low frequency shear waves in soft tissues via radiation force, while the other is used as a vibrometer to measure the resulting shear wave displacements. Results are presented for in situ measurements on formalin fixed tissues from a bottlenose dolphin (tursiops truncatus). [Work supported by ONR.]

3:35

1pBB9. Tissue shear elasticity assesment using time reversal. Gallot Thomas, Catheline Stefan (LGIT,1381 rue de la Piscine, 38041 Grenoble, France), Benech Nicolas, Negreira Carlos, and Brum Javier (LAU, 11400, Montevideo, Uruguay)

Shear waves in a soft solid can be focalized using time-reversal (TR) methods. The wavelength depending on the elasticity of the medium is related to the focal spot width through the Rayleigh critera. This study proposes to quantitatively estimate the mechanical properties inside a 3-D soft solid cavity from one channel TR experiments. Shear waves are created with an external low-frequency ($<$300 Hz) mechanical excitation. The first step of the TR experiment consists in the observation of the forward reverberant elastic field; in a second step the backward field is computed in order to refocus the shear waves. Then the wavelength is deduced from the TR focus spot. Two methods for the observation step have been tested, the first one measuring one component of the vectorial field inside the solid by ultrasonic techniques developed in elastography. This method takes advantage of the ultrasound spatial accuracy and its robustness as regard to shear wave source allows to think about many applications in the medical field, including deep or moving organs. The second one is based on the surface wave measurement by acoustic sensors. This low-cost method could be useful in food industry.

3:50

1pBB10. Time-reversal Rayleigh wave for soft solid characterization. Brum Javier, Benech Nicolas, Negreira Carlos (LAU, Inst. de Fisica, Facultad de Ciencias, Iguá 4225, 11400, Montevideo, Uruguay), Catheline Stefan, and Gallot Thomas (LGIT, 38041 Grenoble, France)

In this work the shear elasticity of soft solids is measured from the surface wave speed estimation. An external source creates mechanical waves in the audible frequency range. Surface waves are detected using acoustic sensors. The surface wave speed estimation is extracted from the complex reverberated elastic field through a time-reversal analysis. These quantitative measurements in a hard and a soft gelatin-based phantom are validated by independent transient elastography estimations. In contrast with other quantitative elasticity assessment methods, one advantage of the present approach is that it is based on a low cost sound technology. Moreover, the robust and simple experiments performed in cheese and soft phantoms, allows thinking of applications in food industry and medicine.

4:05

1pBB11. Towards predicting intracranial pressure using transcranial Doppler and arterial blood pressure data. Pierre D. Mourad (Appl. Phys. Lab., Dept. of Neurosurgery, Box 356470, Univ. of Washington, Seattle, WA 98195, pierre@apl.washington.edu), Caren Marzban, and Michel Kliot (Univ. of Washington, Seattle WA 98195)

Pressure within the cranium (intracranial pressure, or "ICP") represents a vital clinical variable whose assessment—currently via invasive means—and integration into a clinical exam constitutes a necessary step for adequate medical care for those patients with injured brains. In the present work we sought to develop a non-invasive way of predicting this variable and its corollary–cerebral perfusion pressure (CPP), which equals ICP minus arterial blood pressure (ABP). We collected transcranial Doppler (TCD), invasive ICP and ABP data from patients at a variety of hospitals. We developed a series of regression-based statistical algorithms for subsets of those patients sorted by etiology with the goal of predicting ICP and CPP. We could discriminate between high and low values of ICP (above/below 20 mmHg) with sensitivities and specificities generally greater than 70 percent, and predict CPP within $\pm$5 percent, for patients with traumatic brain injury. TCD and invasive ABP data can be translated into useful measures of ICP and CPP. Future work will target use of non-invasive ABP data, automation of TCD data acquisition, and improvement in algorithm performance.

4:20

1pBB12. *In-utero* imaging of developing mouse embryos using a 34-MHz annular array and chirp-coded excitation. Jonathan Mamou, Jeffrey A. Ketterling (F. L. Lizzi Ctr. for Biomedical Eng., Riverside Res. Inst., 156 William St., 9th Fl., 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) is becoming a prevalent modality to image biological tissues with fine resolution. HFU imaging is routinely used for ophthalmologic, dermatological, and small-animal applications. Fine-resolution imaging is obtained at the expense of limited depth-of-field and acoustic penetration depth. In this study, a 34-MHz, five-element annular array was excited using chirp-coded signals and a synthetic-focusing algorithm was used to form images. This imaging approach allowed for the simultaneous increase of depth-of-field and penetration depth. Three-dimensional ultrasound datasets of 12.5-day-old mouse-embryo heads were acquired *in utero* using chirp and conventional excitations. From the synthetically focused images, brain ventricles and embryonic head were segmented and rendered in three dimensions. Significant artifacts such as segmenting ventricles outside the embryonic head were significantly more pronounced on the conventional-imaging reconstruction than on the chirp-based reconstruction. Furthermore, the total ventricle volume obtained with the chirp (6.3 mm^3) was similar to that estimated using a μMRI dataset (i.e., 7.5 mm^3). The conventional method greatly over evaluated the ventricle volume (11.4 mm^3). The results of this study suggest that HFU chirp-annular-array imaging could become a valuable method for medical imaging applications, such as *in utero* developmental biology. [Work supported by NIH Grants EB006509 and EB008606.]

4:35

1pBB13. Micro flexible ultrasound imaging array systems. Michael Lee, Rahul S. Singh, Martin O. Culjat, Elliott R. Brown, Warren S. Grundfest, and Hua Lee (CASIT, UCLA, Los Angeles, CA 90095)

Medical ultrasonography has continued to be an important biomedical imaging modality for its various advantages. Conventional medical ultrasound systems typically utilize a rigid transceiver phase array, with a relatively small transceiver contact area. A micro conformal system increases observation aperture by using a large, flexible transducer array that conforms to curved body surfaces, providing increased angular coverage and 3-D imagery. In addition, the frequency-modulated continuous wave (FMCW) signaling format requires lower peak power than pulse-echo systems and exhibit more favorable receiver noise characteristics. This paper provides an overview of the hardware development of the micro flexible ultrasound array system. The image formation procedure is first presented in the form of pulse-echo imaging modality. Subsequently, image reconstruction algorithms are formulated for the FMCW imaging format. In addition, the equivalence between the pulse-echo and FMCW imaging modes is developed. This enables us to conduct image reconstruction process in both space as well as spatial-frequency domain. The presentation includes theoretical analysis, image formation algorithms, and laboratory experiments. [This research project was supported by the Telemedicine and Advanced Technology Research Center (TATRC), Department of Defense under award numbers W81XWH-07-1-0672 and W81XWH-07-1-0668.]

Session 1pMU

Musical Acoustics and Engineering Acoustics: Microphone Array Techniques in Musical Acoustics

Rolf Bader, Chair
Inst. of Musicology, Univ. of Hamburg, Hamburg 20354, Germany

Invited Papers

1:00

1pMU1. Mapping the vibroacoustic behavior of harp soundboxes. Shira Daltrop, Andrzej Kotlicki, and Chris Waltham (Dept. of Phys. and Astronomy, Univ. of British Columbia, 6224 Agricultural Rd., Vancouver, BC V6T 1Z1, Canada)

Modern harp soundboxes typically consist of a thin soundboard made of an acoustic wood with a definite set of resonances, and a relatively heavy conical or polygonal back in which there are several holes. These holes exist to allow access to the string attachments but also have important acoustical properties. Harps more than 200 years old frequently have small holes in the soundboard itself, and none in the back. The quality of the sound radiation from a harp is determined in large measure by the vibrational behavior of the soundboard and the motion of air in the soundholes. Using an impact hammer, accelerometer and acoustic-velocity probe, we have mapped the velocity of a harp soundboard, and that of the air in the soundholes, in response to excitation at the string attachment points. The relationship between these motions is complex but also crucial in determining the quality of the instrument.

1:30

1pMU2. Real-time guitar radiation sound synthesis of forced string and body eigenfrequency vibrations using microphone array techniques. Rolf Bader (Inst. of Musicology, Univ. of Hamburg, Neue Rabenstr. 13, 20354 Hamburg, Germany, R_Bader@t-online.de)

A guitar top plate vibration is investigated for different frequency regions taking string driving point dependencies into consideration using a microphone array of 128 microphones and back-propagating the recording to the top plate positions. These radiation patterns are integrated in space at arbitrary virtual microphone positions. A guitar tone or piece is then synthesized using these radiation patterns where the microphone positions can be changed as in a real recording situation. The synthesis process takes the pressure and phase relations of more than 100 positions on the top plate and the sound hole into consideration for two cases of waves, the traveling waves of forced top plate vibration of the string frequencies and the eigenfrequencies of the guitar body appearing mostly during the initial transient phase. These eigenfrequencies are integrated using the much more simple mode shapes of the top plate eigenmodes. The amplitude behavior of these frequencies, appearing and decaying during the initial transient, is calculated by integrating the energy which is supplied by the string acting with a force on the top plate as an impulse train.

2:00

1pMU3. Acoustic holography measurements in comparison with a three-dimensional finite element method modulation of the body of a classical guitar. Jan Richter (Inst. of Musicology, Univ. of Hamburg, Neue Rabenstrasse 13, 20354 Hamburg, Germany, janrichter81@web.de)

The body of a classical guitar (consiting of the rib, back plate, top plate with bridge and fan bracing, soundhole and wooden block) which had been performed by 3-D finite element method (FEM) was presented at the 151st meeting of the ASA (session 3aMU3 in Providence, RI). As a submittal, a real guitar was used whose geometrical parameters had precisely been modeled using a computer. The resulting computation has subsequently been matched to the measurements of an acoustical holography taken from the same guitar body. This time the aim was to demonstrate how closely the FEM model resembles the "real" sound radiation of the guitar. Acoustic holography graphically displays the characteristics of sound radiation from a given body. Again the mode-shapes are of interest, but this time the characteristics of the sound in relation to the frequencies are of greater importance. They demonstrate how realistic the modulation is compared to the actual acoustics of the guitar.

2:30—2:45 Break

2:45

1pMU4. Realtime virtual banjo model and measurements using a microphone array. Florian Pfeifle (Musik-wissenschaftliches Institut, Univ. of Hamburg, Alte Raabenstrasse 13, 20345 Hamburg, Germany, Florian.Pfeifle@haw-hamburg.de)

A model of a full-scale banjo was programmed in MATLAB and C using the finite differences method. Because of the highly complex algorithm, the model could not be put into practice in real time. A similar model was then written in VHDL, a hardware description language and implemented on a FPGA board. This hardware allows highly parallelized computation at highest clock rates, thereby, a realistic banjo model could be computed in real time. This virtual banjo could now be used to alter instrument-specific parameters in real-time. A large set of tests was conducted ranging from simple calibration of the most realistic parameters to the research of nonlinear effects on the membrane of the banjo. The findings where compared to the measurements of a real banjo done with a microphone array consisting of 128 microphones.

3:15

1pMU5. Acoustic and vibrometry analysis of beating in a large Balinese gamelan gong. David W. Krueger, Kent L. Gee (Dept. of Phys. and Astron., Brigham Young Univ., N283 ESC, Provo, UT 84602, dvdkrueger@gmail.com), Jeremy Grimshaw (Brigham Young Univ., Provo, UT 84602), Dorothy Manuel (Spelman College, Atlanta, GA 30314), and Molly E. Jones (Univ. of Michigan, Ann Arbor, MI 48109)

A large Balinese gamelan gong, the gong ageng wadon, is similar to other Indonesian gongs in that it is tonal. A previously undocumented phenomenon is the fact that this gong produces distinct beating, which is essential to Balinese gamelan music. In this study of Brigham Young University's gong ageng wadon, acoustical and vibrometry measurements were performed to analyze the source of the beating. Scanning laser Doppler vibrometer results show a beat frequency of about 3 Hz is produced near 150 Hz by closely spaced structural modes. A slightly slower beat frequency (around 2.5 Hz) is also produced at about 120 Hz. The difference in this case is that there is only one structural mode measured near this frequency. It appears that the acoustic radiation from this mode beats with a harmonic of the radiation from a lower structural mode and that this harmonic is caused by nonlinear acoustical phenomena.

3:30

1pMU6. Dampening vibration modes of the snare drum batter head. Barry Larkin (Dept. of Music, Iowa State Univ., 2825 W. Lincoln Way, Ames, IA 50014, blarkin@iastate.edu) and Andy Morrison (Northwestern Univ., Chicago, IL)

Percussionists have always had to contend with an undesirable ringing sound while performing on the snare drum. It is usually referred to as the "edge ring." A common method to eliminate this sound comes from placing some type of dampening material on the edge of the drum head. Manufacturers of drums have provided many ways of dealing with this problem including internal dampening devices, customized drums heads, or material designed to be placed on the drum head. Using electronic television holography, it was revealed the source of this "ring" to be the third mode of vibration that produces a pitch approximately one octave and a half step above the fundamental. During this presentation, laser holographic images from the fundamental to the 16th mode will be displayed. In addition, new and recent holography work will show the effect of dampening devices on the third and other modes. "I certify that I have complied with ASA ethical principles" Telephone number: 515.294.2818, FAX: 515.294.6409. Send Notice to Barry Larkin.

3:45

1pMU7. Normal modes of a circular drum head under nonuniform applied tension: Mode splitting with respect to perturbation symmetry. Randy Worland (Dept. of Phys., Univ. of Puget Sound, 1500 N. Warner, Tacoma, WA 98416, worland@ups.edu)

The effect of nonuniform tension on the normal modes of single headed musical drums is examined. Using the common (m,n) designation, ideal circular membrane modes with $m>0$ are doubly degenerate. This degeneracy may be lifted by a perturbation to the circular symmetry. Of practical interest to drummers are perturbations caused by nonuniform tension applied at the rim of the drum in the tuning process, leading in some cases to audible frequency splitting. The symmetry of the perturbation determines which of the various two-fold degenerate modes will split, and to what degree the two resulting frequencies will differ. A selection rule taken from group theory is used, along with principles of perturbation theory, to describe a general model of the frequency splitting of drum modes under the influence of nonuniform applied tension. This model is compared with results of finite element modeling of the membrane normal modes. Experimental results obtained using electronic speckle pattern interferometry (ESPI) are also compared with the model.

4:00

1pMU8. Importance of visual cues in networked music performances. Jonas Braasch (Sch. of Archit., Rensselaer Polytech. Inst., 110 8th St., Troy, NY 12180, braasj@rpi.edu)

Although music is predominantly an acoustic event, the importance of visual cues should not be underestimated. In this paper, a number of observations will be presented that were derived both from laboratory experiments and ongoing music collaborations over the Internet. Different requirements exist for networked music performances than for telematic speech conferences for which most Internet-based telecommunication systems have been developed. For example, audio/visual synchronicity is essential for speech applications, but it was often found advantageous to run the audio ahead of the video in networked music performances to minimize the audio delay, which needs to be below a critical value if the remote ensembles perform together rhythmically. Other aspects of speech and music applications are very similar, for example, the need to exchange documents, which are in the case of music predominantly music scores. In the common telepresence approach, it is typically asked how realistic the recreated environment has to be for satisfying results. In this paper, an alternative, more abstract approach based on virtual functionality will be discussed that also includes the integration of music scores into the virtual environment. [Work supported by NSF Grant 0757454.]

Session 1pNS

Noise, Animal Bioacoustics, and ASA Committee on Standards: Bioacoustic Metrics and the Impact of Noise on the Natural Environment

Michael Stocker, Chair
Ocean Conservation Research, P.O. Box 559, Lagunitas, CA 94938

Chair's Introduction—1:25

Invited Papers

1:30

1pNS1. Predicting the effects of masking noise on communication distance in birds. Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742), Marjorie Leek (Portland, OR 97225), and Ed West (West Ecosystems Anal., Inc., Davis, CA 95616)

Anthropogenic noise (e.g., highway noise) that exceeds natural ambient sound levels potentially can mask important acoustic signals used by birds during communication. Signals degraded by masking can in turn adversely affect critical behaviors such as mate attraction, territorial defense, parent-offspring communication, predator avoidance, and prey detection. The amount of masking depends not only on the physical characteristics of the noise (intensity, frequency, temporal pattern, etc.) but also the auditory capacity of the species, the acoustic structure of their vocal signals, the proximity of the birds to the noise source and each other, and the sound attenuation properties of the habitat. Here we present a quantitative model for predicting effective communication distances between birds subject to different levels and types of anthropogenic noise. Cross-species analysis shows that communication distances are largely context /environment dependent and species specific. Extrapolation of noise impact distance estimates even within groups of similar species should be exercised with caution. This model should be useful in environmental impact analysis of anthropogenic noise on birds and other wildlife, particularly species of conservation concern.

1:50

1pNS2. Metrics for characterizing the sources of ocean anthropogenic noise. John A. Hildebrand (Scripps Inst. Oceanogr., Univ. of California, San Diego, La Jolla, CA 92092-0205)

Decibels are the standard shorthand for describing acoustic intensity and sound pressure level, but may lead to misunderstanding when applied as bioacoustic metrics. Acoustic power and source transmission energy are alternate metrics with intuitive appeal. Acoustic power, calculated from the acoustic intensity multiplied by the emitted solid angle, yields units of watts. Likewise, the energy per source transmission, given by multiplying acoustic power by the duration of the transmission, yields units of joules. For continuous (or quasicontinuous) signals, the standard procedure is to measure the root-mean-square (rms) of the signal. However, this presents problems for short duration signals where the duration of the signal being measured is an important parameter. In these cases it may be more appropriate to measure the peak-to-peak signal, rather than rms. Bandwidth is another important component of how the signal is described, typically in a narrow-band for ambient noise and broad-band for discrete sources. The characteristics of ocean anthropogenic noise sources in terms of these metrics will be discussed.

2:10

1pNS3. Marine mammal noise exposure criteria: Initial scientific recommendations. Brandon L. Southall (NOAA/NMFS Ocean Acoust. Prog., 1315 E-W Hwy. #12539, Silver Spring, MD 20910), Ann E. Bowles (Hubbs-SeaWorld Res. Inst., San Diego, CA 92109), William T. Ellison (Marine Acoust., Inc., Middletown, RI 02842), James J. Finneran (Space and Naval Warfare Sys. Ctr., San Diego, CA 92152), Roger L. Gentry (ProSci. Consulting, LLC, Dickerson, MD 20842), Charles R. Greene, Jr. (Greeneridge Sci. , Inc., Santa Barbara, CA 93110), David Kastak (U.C. Santa Cruz, Santa Cruz, CA 95060), Darlene R. Ketten (Woods Hole Oceanog. Inst., Woods Hole, MA 02543), James H. Miller (Univ. of Rhode Island, Narragansett, RI 02882), Paul E. Nachtigall (Hawaii Inst. of Marine Biol., Kaneohe, HI 96744), W. John Richardson (LGL Ltd., King City, ON, Canada), Jeanette A. Thomas (Western Illinois Univ.-Quad Cities, Moline, IL 61265), and Peter L. Tyack (Woods Hole Oceanog. Inst., Woods Hole, MA 02543)

An expert panel reviewed the expanding literature on marine mammal (cetacean and pinniped) auditory and behavioral responses to sound exposure to develop comprehensive, scientifically based noise exposure criteria [Aquatic Mammals **33**(4)]. They used precautionary extrapolation procedures to predict exposure levels above which adverse effects (both physical and behavioral) could be expected. Due to the paucity of data on long-term exposures, criteria were developed for single exposure events only. Marine mammals were broken into functional hearing groups. Exposure types were lumped into three broad classes (single pulses, multiple pulses, and nonpulses). Levels estimated to induce permanent noise-induced hearing loss were determined for each of 15 sound type/animal group combinations. For example, injury criteria for pinnipeds in water exposed to multiple pulses were 186 dB re 1 μPa^2 -s (weighted SEL) and 218 dB_{pk} re 1 μPa (unweighted peak SPL). Discrete behavioral disturbance thresholds could only be determined for exposure to single pulses. For other exposures, available data on behavioral responses were ranked by severity and significance. This severity scaling and the resulting conclusions will be described. New research required to improve criteria and to assess cumulative and ecosystem-level effects will also be considered, along with current policy and/or regulatory applications.

2:30

1pNS4. The importance of incorporating signal characteristics in the evaluation of noise exposure impacts on marine life. Michael Stocker (Ocean Conservation Res., P.O. Box 559, Lagunitas, CA 94938)

Auditory thresholds are used to establish mitigation guidelines for anthropogenic noise exposure on marine animals. These thresholds are determined using either sinusoidal signals at specific frequencies or band limited, sinusoidal-derived Gaussian noise. Given that the preponderance of naturally occurring noise in the marine environment is sinusoidal, marine animals may have lower thresholds and, thus, lower tolerance to non-sinusoidal noise. Fast rise time impulse noise, continuous non-sinusoidal noise, or a combination of these characteristics may induce biological responses at lower levels than sinusoidal noise with an equivalent power density. The author proposes a metric to evaluate and express signal characteristics as a component of determining noise exposure impacts on marine animals.

2:45

1pNS5. Role of the National Marine Fisheries Service in underwater acoustics: Integration of science and policy. Amy R. Scholik-Schlomer, Shane Guan, and Jolie Harrison (Natl. Marine Fisheries Service, Office of Protected Resources, 1315 East-West Hwy., Silver Spring, MD 20910, amy.scholik@noaa.gov)

The National Marine Fisheries Service, Office of Protected Resources (NMFS OPR) is responsible for protecting marine mammals and other endangered marine species and works to conserve and recover these species under the Marine Mammal Protection Act and Endangered Species Act. Accurately assessing the impacts of anthropogenic underwater noise on the marine environment and implementing appropriate measures to reduce such impacts is one of its primary focuses. To do this successfully requires the integration of science with policy. This integration combines a complex set of factors. NMFS OPR has jurisdictions over a variety of marine species, considers a wide range of potential impacts, and regards a diverse assortment sound sources. Additionally, it works collaboratively with other federal, state, and local government agencies, academia, nongovernmental organizations, and industry, as well as solicits public input on the actions it regulates. NMFS OPR uses and relies on scientific data to make informed decisions on numerous issues relating to anthropogenic underwater noise, including developing appropriate acoustic thresholds for assessing impacts and best protecting marine species via the establishment of practicable mitigation and monitoring protocols. To most effectively achieve these goals requires continual interaction with and input from those in the scientific community.

3:00

1pNS6. Methods for computing the impact of sonar on the marine environment. Martin Siderius (Dept. of Elect. and Comput. Eng., Portland State Univ., 1900 SW 4th Ave., Portland, OR 97201, siderius@pdx.edu), Dorian Houser (BIOMIMETICA, Santee, CA 92071), Daniel Hernandez, and Michael Porter (HLS Res. Inc., La Jolla, CA 92037)

Federal agencies are required to estimate the impact to the environment resulting from major federal actions. The U.S. Navy is required to prepare environmental impact statements (EISs) for a number of its training ranges, in part, due to the controversial nature of the sonar systems used at these sites. Over the past several years these EISs have been prepared by both government agencies and by contractors. This multi-institution approach has led to different methodologies being used to estimate impacts on marine mammals. Different impact estimates and confusion as to the real risk posed by a particular sonar system has resulted. As part of the Office of Naval Research program on the effects of sound on the marine environment (ESME), we have been assembling state-of-the-art modeling software to provide the Navy and the Public the best risk assessment tools currently available. One of the challenges to this effort is determining which approaches provide the best estimates of risk and which are practical to implement. Related to this is the development of an objective set of metrics to make such evaluations. In this presentation we will describe the risk modeling methodologies and the metrics being used for analysis in the ESME model.

3:15—3:30 Break

3:30

1pNS7. Ambient noise properties in a region of Puget Sound. David Dall'Osto and Peter H. Dahl (Appl. Phys. Lab. and Dept. of Mech. Eng., Univ. of Washington, Seattle, WA 98105)

Ambient noise data collected during August 2007 at a site near the Smith Island Marine Sanctuary in North Puget Sound, Washington, depth 100 m, is discussed. This marine sanctuary is representative of a marine mammal habitat, where adverse impacts of noise are of concern. Properties of the ambient noise environment are inspected relating to frequency content and vertical directionality. Wind speed and wave spectral data for correlations were provided by the Smith Island weather station and a nearby NOAA buoy. Two co-located systems were used to estimate spectral levels of the ambient noise environment. The directional information is provided by one of these systems consisting of a pair of omni-directional elements vertically separated by 15 cm. Candidate models for noise directionality are tested against the array gain for this simple system evaluated as a function of frequency, giving an approximate picture of the noise density as a function of vertical arrival angle at particular frequencies of interest. [Research supported by Washington Sea Grant.]

3:45

1pNS8. Seismic airgun sounds recorded in the North Atlantic. Sharon L. Nieukirk, David K. Mellinger (Coop. Inst. for Marine Resour. Stud., Hatfield Marine Sci. Ctr., Oregon State Univ., 2030 S. Marine Sci. Dr., Newport, OR 97365, sharon.nieukirk@oregonstate.edu), Sue E. Moore (Univ. of Washington, Seattle, WA 98105), Karolin Klinck, Sara L. Heimlich, and Holger Klinck (Oregon State Univ., Newport, OR 97365)

Between 1999 and 2008, a consortium of international investigators deployed autonomous hydrophones to monitor seismic activity along the Mid-Atlantic Ridge. Although this experiment was designed to monitor low-frequency sounds from earthquakes, other natural and man-made sounds were also recorded. The periodic impulses produced by airguns from exploration activities around the Atlantic basin were one of the dominant signals detected in this dataset. Sounds associated with seismic exploration, particularly sounds from airgun arrays, have garnered interest recently due to concerns regarding the potential impact on marine mammals. In a previous publication, we reported that in 1999–2001 airgun sounds were recorded year-round on all hydrophones, but airgun activity peaked in the summer months. For this study, we used an automatic algorithm to detect airgun sounds in the 2001–2008 portion of the dataset. Raw detections were then visually examined by an experienced analyst to verify the accuracy of the detections. When airgun impulses were simultaneously recorded on four or more hydrophones, the location of the seismic exploration vessel was estimated. Results from this dataset will be presented in light of current knowledge of marine mammal distribution and migration patterns in the Atlantic.

4:00

1pNS9. Assessment of bubble curtain attenuator techniques for underwater pile driving noise mitigation. Michael A. Slater (SAIC Maritime Technologies, 26279 Twelve Trees Ln., Ste. B, Poulsbo, WA 98370)

Public concern for prudent management of marine resources due to anthropogenic activity is large and growing. In particular, the potential for causing negative impacts to marine species including salmonids and mammals is sufficient motivation to consider all reasonable mitigation strategies to minimize these effects. An impact pile driver can introduce extremely high sound pressure levels into the environment, and a bubble curtain attenuator is one means to reduce noise propagated from each driven pile by changing the impedance along the primary path of propagation. Available literature describing the observed effectiveness of reductions measured dur-

ing *in situ* pile driving monitoring while using a bubble curtain range from completely ineffective to substantially effective. As a result, opposing views both seek acoustic evidence to support the need to use, or not use, bubble curtain attenuators during impact pile driving activities. This paper provides a description of the current state of knowledge on the use of bubble curtain attenuators, outlines on-going activities to address the issue, identifies limitations to the knowledge base, and suggests recommendations to close the information gap.

4:15

1pNS10. Measurements of industrial noise and vibration in artificial polar bear (*Ursus maritimus*) dens, Flaxman Island, Alaska. Alexander O. MacGillivray, David E. Hannay, and Craig J. Perham (JASCO Res. Ltd., Markham St., Ste. 2101-4464, Victoria, BC V8Z 7X8, Canada, alex@jasco.com)

Noise and ground vibration generated by environmental remediation activities were measured in man-made polar bear (*Ursus maritimus*) dens at Flaxman Island, Alaska in Mar. 2002. The remediation activities involved use of heavy excavation equipment, ground vehicles, and blasting. Comparison of sound levels measured inside and outside the dens permitted estimation of the sound-insulating properties of the dens. Vibration data were acquired from sensors placed in the tundra and in the snow of the den floors. Measurements of noise and vibration were made for the following vehicles: front-end loader, trimmer, gravel hauler, fuel truck, pickup truck, Hägglunds tracked vehicle, and a Tucker Sno-Cat. A single 8.7 kg blast event at 19 m underground, which was used to cut a well pipe, was recorded in the dens at distances from 500 to 1300 m. Additional noise and ground vibration data were obtained for Bell 212 and Bell 206 helicopters performing maneuvers above the dens. The man-made dens were found to have good sound-insulating properties. Sound levels inside the dens were lower than directly outside the dens by 25 dB at 50 Hz, and by 40 dB at 1000 Hz.

4:30

1pNS11. Mediterranean fin whale migration movements altered by seismic exploration noise. Manuel Castellote (Oceanografic of Valencia and Psychobiology Dept., Univ. Complutense of Madrid, Spain, mcastellote@oceanografic.org), Christopher W. Clark (Cornell Univ., New York), Fernando Colmenares (Univ. Complutense of Madrid, Spain), and José Antonio Esteban (Oceanografic of Valencia, Spain)

Seafloor recorders were deployed along the Spanish Mediterranean coast during 2006–2009 to further contribute to the little knowledge regarding Mediterranean fin whale (*Balaenoptera physalus*) movement patterns. Typical long patterned sequences of 20-Hz pulses were identified in different periods. Detection results suggest that at least part of the Mediterranean fin whales aggregating in the Ligurian-Corsican-Provençal basin during summer migrate through French and Spanish waters toward southern Mediterranean regions, including the south of Spain. Continuous airgun explosions from a seismic survey were detected during 8–17th December 2006 in the south of Spain. A concurrent fin whale reaction in acoustic behaviour was documented: All singing whales moved away from the airgun area and out of the detection area during the following 20 days, even after the airgun noise ceased on the tenth day, and remained weak after the 20-day period. Twenty-hertz pulse bearings for this period revealed that calling whales were moving through the sensors detection area and not that whale acoustic behaviour was reduced by the seismic activity. We therefore conclude that the seismic airgun activity caused fin whales to avoid their potential winter ground for a period of time that extended well beyond the duration of the seismic airgun array activity.

4:45

1pNS12. Acoustic characterization of whale watching boats in the Azores. Cláudia Oliveira, João Gonçalves, Maria João Cruz, Mónica A. Silva, Rui Prieto (Dept. de Oceanografia e Pescas, Centro do IMAR da Univ. dos Aores, Horta, Portugal), Marc Lammers (Univ. of Hawaii, Kaneohe, HI), and Ricardo S. Santos (Centro do IMAR da Univ. dos Aores, Horta, Portugal)

The whale-watching activity has been growing in the Azores during the last 15 years. To contribute to the sustainability and management of the activity it is crucial to monitor its influence on cetacean populations. This work aimed to determine the sound produced by whale-watching boats, in terms of frequencies and sound pressure levels (SPLs). Different types of boats were recorded between May and Sept. 2007 around the islands of Faial, Pico, and S. Miguel. Recordings were made with a hydrophone from a drifting platform. The whale-watching boats carried out several tracks with different speeds and changed gear (forward/reverse). SPL values varied between 74 and 116 dB re 1 μPa (1/3 octave band level) for both outboard and inboard engines, without any obvious difference between the two. One inboard catamaran produced 106 dB re 1 μPa at 50 kHz (18 knots, 50 m distance), which is higher than all the other SPLs, at the same frequency. Forward/reverse maneuvers in inboard engines had lower frequencies but a longer duration than in outboard boats. Frequency range produced by the Azorean whale-watching boats is wide and the way boats are maneuvered is important to minimize their impact on cetaceans. [Work was supported by Direcção Regional do Turismo do Governo Regional dos Açores and Secretaria Regional do Ambiente e do Mar do Governo Regional dos Açores.]

Session 1pPA

Physical Acoustics: General Physical Acoustics

James M. Sabatier, Chair
National Ctr. for Physical Acoustics, Univ. of Mississippi, University, MS 38677

Contributed Papers

1:30

1pPA1. Effects of soil water potential and moisture content on sound speed. Zhiqu Lu and James Sabatier (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677)

In order to monitor and understand the temporal variations of sound speed due to the changes of soil physical properties under natural conditions, a long-term field soil survey has been conducted. Measurements of sound speed, soil temperature, soil moisture content, and water potential were performed continuously, along with the measurements of surface temperature and precipitation over a period of 2 years. The effects of water potential, moisture content, soil temperature, and depth on sound speed were assessed. Analysis of the data shows that there is a power law relationship between the sound speed and water potential. It is also found that the water potential is the dominant influence on the sound speed, whereas the moisture content and temperature have relatively minor impacts. Brutsaert and Luthin's theory was employed to calculate the sound speed as a function of the effective stress. The theoretical predictions were compared with the experimental data and they are in good agreement. The study suggests that sound speed measurement could be used as a new and effective tool for water potential measurement.

1:45

1pPA2. Multichannel analysis of surface waves method for shallow soil profile exploration. Zhiqu Lu and James Sabatier (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677)

In order to explore shallow soil properties in a noninvasive manner, a multichannel analysis of surface waves method was employed to obtain soil shear wave velocity profile up to a couple of meters. In the study, an electromechanical shaker operating in a frequency-sweeping mode was placed on the ground to excite Rayleigh waves propagating along the surface. A laser Doppler vibrometer (LDV) was used as a noncontact receiver to detect the surface vibration. The LDV was mounted on an aluminum frame and driven by a stepper motor to move along a straight line. Multichannel surface vibrations were obtained and analyzed to determine the dispersive curve; i.e., the Rayleigh wave velocity as a function of frequency. An inversion algorithm was used to transform the dispersion curve into the shear wave velocity profile. This technique can be developed for the applications of levee and dam evaluation as well as shallow soil properties imaging and characterization.

2:00

1pPA3. Wideband multidomain ultrasonic transducer. Andriy Nadtochiy and Igor Ostrovskii (Dept. of Phys., Univ. of Mississippi, University, MS 38677)

Wideband multidomain ultrasonic transducer (WMUT) based on a ZX-cut aperiodically poled LiTaO3 is considered and demonstrated. For a wide band operation inside a frequency range of 40–80 MHz, the domain sizes within a multidomain array vary from 25–50 μ, and the total number of ferroelectric domains N varies from 25–266. The partial-waves method computation yields a strong effect of a narrow acoustic signal formation by a WMUT. For instance, an ultrasonic signal in water is about 10 times shorter than input rf chirp burst applied to the WMUT. As the number N increases, more frequencies in the acoustic signal are present. Experimentally it is demonstrated a wideband delay line for a chirp signal within the frequency range of 32–64 MHz. The delay line is fabricated in a 500-μ thick LiTaO3 wafer, and has two WMUTs with variable domain sizes from 25–50 μ. Ultrasound is generated by an input WMUT having $N=14$, and then detected with output WMUT having $N=134$. The experimental measurements are in agreement with theoretical computations. [Work supported by UM.]

2:15

1pPA4. Modeling acoustic response of permeability variation in deepwater wells using spectral method. Florian Karpfinger (Dept. of Expl. Geophy., Curtin Univ. of Tech., P.O. Box U1987, Perth, WA 6845, Australia), Boris Gurevich (Curtin Univ. of Tech., Perth, WA 6845 Australia and CISIRO Petroleum, Benteley, WA), and Andrey Bakulin (WesternGeco, Houston, TX 77042)

Petroleum production in deepwater oil fields often hinges on the ability to safely complete and effectively draw down a small number of very challenging wells. The success of this operation requires regular completion monitoring, which is most effectively achieved using real-time surveillance tools. Real-time completion monitoring with acoustic waves (RTCM) has great potential for diagnosing problems in sand-screened deepwater completions. RTCM uses tube waves to detect permeability changes and passive noises to characterize perforation flow. Interaction of a single tube wave with permeable formations in open boreholes is well explained by Biot's theory of poroelasticity. However, experimental studies in laboratory models of sand-screened completions reveal existence of two tube wave modes: fast and slow, supported by completion and sand screen, respectively. Development of quantitative monitoring methodology of acoustic surveillance requires better understanding on how signatures of fast and slow tube waves depend on completion properties. To this end, we simulate two tube dispersion and attenuation by examining the solutions of Biot's equations of poroelasticity in cylindrical structures using a spectral method. The results show dependence of the velocity and attenuation of both tube waves on completion permeability. These modeling results are in a qualitative agreement with laboratory experiments.

2:30

1pPA5. Spectral method for modeling waves in cylindrical poroelastic structures. Florian Karpfinger (Dept. of Expl. Geophys., Curtin Univ. of Technol., P.O. Box U1987, Perth, WA 6845, Australia), Boris Gurevich (Curtin Univ. of Technol., Perth, WA 6845, Australia and CISIRO Petroleum, Bentley, WA), Andrey Bakulin (WesternGeco, Houston, TX 77042), and Henri-Pierre Valero (Schlumberger–Doll Res., Cambridge, MA 02141)

The modeling of propagating modes in poroelastic fluid-saturated cylindrical structures is important for the interpretation of acoustic and seismic measurements in hydrocarbon wells as well as laboratory measurements on core samples. In order to obtain the velocities of different modes the roots of the appropriate dispersion equation have to be found. This is a very difficult task especially when some layers are represented by poroelastic materials with frequency-dependent attenuation. A new technique is presented for modeling those modes, which adopts a spectral method to discretize the underlying partial differential equations using Chebyshev differentiation matrices. The corresponding system of linear equations can then be solved as a generalized eigenvalue problem. This means that, for a given frequency, the eigenvalues correspond to the slownesses of different modes. This approach is implemented for an arbitrary number of cylindrical fluid and solid

layers as well as poroelastic layers. In addition to the dispersion, the attenuation curves are computed as well as radial profiles of displacements and stresses for various frequencies. The resulting approach can be used to model wave propagations in open and completed boreholes surrounded by permeable formations. This can pave the way to invert for formation and completion permeability from acoustic measurements.

2:45

1pPA6. The penetration of sound into an extended reaction ground due to an arbitrarily oriented dipole. Kai Ming Li and Sheng Liu (Ray W. Herrick Labs., Purdue Univ., 140 S. Martin Jischke Dr., W. Lafayette, IN 47907-2031)

An asymptotic solution is derived for the sound fields below an extended reaction ground due to an arbitrarily oriented dipole placed near the air /ground interface. The derived approximate solutions are compared with accurate numerical schemes based on the fast field formulation and the direct numerical integration of the Fourier integral. It has been shown that the sound field above the ground can be computed accurately by using the concept of effective impedance [K. M. Li *et al.*, J. Acoust. Soc. Am. **104**, 679–685 (1998)]. The effective impedance approach is used to derive the asymptotic formula for predicting the penetration of sound into the extended reaction ground. However, there are apparent discrepancies in the predicted sound fields according to the analytical solution and the accurate numerical schemes. A modified analytical formula is then developed that leads to more accurate numerical results. The resulting analytical formula due to an airborne acoustic excitation may be used as a basis for a future study to compute the particle displacements induced by an elastic object buried in a rigid-frame porous medium.

3:00—3:15 Break

3:15

1pPA7. Energy flux imaging method for time reversal in elastic media. Brian E. Anderson, Robert A. Guyer, Timothy J. Ulrich, Pierre-Yves Le Bas, Carene Larmat, Michele Griffa, and Paul A. Johnson (Geophys. Group EES-17, Los Alamos Natl. Lab., Los Alamos, NM 87545)

An energy flux imaging method for use in locating sources of wave energy during the back propagation stage of the time reversal process will be presented. During the back propagation phase of an ideal time reversal experiment, wave energy coalesces from all angles of incidence to recreate the source event; after the recreation, wave energy diverges in every direction. An energy flux imaging method based on this convergence/divergence behavior has been developed. The energy flux imaging method yields better spatial resolution in source reconstruction than is possible with traditional imaging methods. [This work is supported by Institutional Support (LDRD) at Los Alamos National Laboratory.]

3:30

1pPA8. Impedance tube measurements of gravel at frequencies between 25 and 200 Hertz. Michael J. White (U.S. Army Engineer Res. and Develop. Ctr., PO Box 9005, Champaign, IL 61826), George W. Swenson, Jr., and Michael L. Oelze (Univ. of Illinois, Urbana, IL 61801)

The sound absorption properties for a few materials have been examined for noise control, yet few materials are as durable and as cheap as gravel. We have measured the attenuation coefficient and wavelength in crushed limestone and pea gravel using an impedance tube. The tube is made of heavy steel, has 0.76-m inner diameter, height 7.3 m, and is fitted with 31 microphones flush-mounted on a line along the inside wall, using 0.2-m spacings. The tube can be filled with a variable depth (0–7 m) of test material, to be supported by either a perforated or solid rigid end plate. Over the frequencies 25–200 Hz, sound speed in 12-mm crushed limestone and 6-mm unsifted pea gravel increased from 160 to 200 m/s. Strongly graded 6-mm pea gravel was poorly dispersive, with sound speed averaging 210 m/s. Over the same frequency interval, the absorption coefficient increased from 0.4 to 1.1 Np/m in coarsely-graded pea gravel. The same positive trend, but successively lower attenuation, was found for crushed limestone and strongly-graded pea gravel. So far, the impedance has been more difficult to determine. Next, we plan to incorporate measurements of complex pressures in an attempt to improve the estimates.

3:45

1pPA9. Penetration depth for measurements with ultrasonic diffraction grating spectroscopy. Margaret S. Greenwood (Pacific Northwest Natl. Lab., P.O. Box 999, Richland, WA 99352, margaret.greenwood@pnl.gov)

The current status of ultrasonic diffraction grating spectroscopy was described at the Acoustics 08 Conference in Paris. An important question remains: What is the penetration depth of the ultrasound for such measurements? Although the spectral order $m=1$ at 90 deg is of primary interest, much insight can be gained by investigating a range of angles. The first step was to measure the amplitude of the $m=1$ transmitted beam in the liquid as a function of the transmitted angle. Data were obtained by sending a toneburst signal to the send transducer and measuring the response of a receive transducer immersed in the liquid. The angle of the receive transducer was varied in half-degree steps from about 50 to 90 deg. The data were compared with a formulation for the interference of N wavelets from N grooves and showed very good agreement. In addition, an experimental technique was developed to measure the penetration of ultrasound into the liquid. Experimental parameters that affect this penetration depth were also investigated. The results of these measurements will be discussed and compared with the theory.

4:00

1pPA10. Self-calibrating measurements of the density and velocity of sound from the reflection of ultrasound at two solid-liquid interfaces. Margaret S. Greenwood (Pacific Northwest Natl. Lab, P.O. Box 999, Richland, WA 99352, margaret.greenwood@pnl.gov)

This report describes a method to measure the density of a liquid and velocity of sound by observing two reflections of ultrasound from a solid-liquid interface: one for perpendicular incidence of a longitudinal wave and the second for incidence of a shear wave at an angle of 45 deg. In previous research [J. Fluids Eng. **126**, 189 (2004)], transducers were mounted on the outside of a stainless steel pipeline, 6 mm thick. Multiple reflections within the wall yielded the acoustic impedance. This method is self-calibrating because each echo is affected in the same way by any change in transducer voltage. The velocity of sound was obtained from the time-of-flight (TOF) across the pipeline. The current research replaces the TOF measurement by using a triangular stainless steel wedge (45, 90, 45 deg). A 1-MHz shear wave transducer is mounted on one short side of the wedge, with the longer side immersed in the liquid. Multiple reflections from the opposite side of the wedge are obtained. The two reflection coefficients can be used to solve for the density and velocity of sound. Greater precision can be obtained by using fused quartz instead of stainless steel.

4:15

1pPA11. Numerical simulation of fluctuated ultrasound wave propagation through random medium. Hideyuki Nomura and Tomoo Kamakura (Dept. of Electron. Eng., Univ. of Electro-Commun., 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan, nomu@ee.uec.ac.jp)

Multiply scattered ultrasound wave propagation in a random medium is numerically simulated to consider the relationship between the physical properties of the random medium and wave fluctuations propagating through that medium. The medium is assumed to be randomly distributed scatters, which are characterized by the means and variances of number, size, density, and volume elasticity of scatters in a host medium. The ultrasound propagation is simulated using a finite-difference time-domain method. To evaluate the fluctuation of propagating ultrasound wave, the ratio of coherent intensity to incoherent and the ratio of the average wave energy to difference wave are statistically estimated. Numerical experiments show that the ratios of intensities and average and difference wave energies decrease with increasing the randomness and number density (scattering cross-sectional area) of scatters. This result suggests that the statistical estimation of wave

pulse propagation is useful for measurements of physical properties of distributed particles in a solution. It is our expectation that the proposed numerical model will provide fundamental understandings on multiple scattering of ultrasound in a random medium.

4:30

1pPA12. Modeling focused ultrasonic beams through liquid-solid interfaces: Applications to diffuse scattering. 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)

When a focused ultrasonic beam interacts with a liquid-solid interface, the beam profile changes due to the acoustic impedance mismatch between the two materials at the interface. A normal incidence focused beam, for planar or curved liquid-solid interfaces, can be modeled with good accuracy using several well established models such as single Gaussian, a multi-Gaussian, Gauss–Hermite, or Bessel beams, which are often used with a paraxial approximation as well. For oblique angles of incidence, these theories break down and the displacements within the solid are not modeled accurately. If such models are applied to diffuse backscatter experiments, accurate extraction of microstructural parameters becomes difficult. In this presentation, ray models of ultrasonic beams are examined with respect to focusing through a liquid-solid interface at oblique incidence as a function of space and time. The integration of such a beam model within the framework of a diffuse ultrasonic scattering theory is discussed, and the theory is used to interpret diffuse ultrasonic scattering experimental data for which focused beams are used at oblique incidence on polycrystalline metals. These results are anticipated to impact ultrasonic nondestructive evaluation of heterogeneous media.

MONDAY AFTERNOON, 18 MAY 2009 GRAND BALLROOM I, 1:30 TO 4:30 P.M.

Session 1pPP

Psychological and Physiological Acoustics: Psychophysics, Modeling, and Signal Processing (Poster Session)

Walt Jesteadt, Chair

Boys Town National Research Hospital, 555 North 30th St., Omaha, NE 68131

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 even-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

1pPP1. The influence of masker fringe on sound localization in noise. Brian D. Simpson (Air Force Res. Lab., 2610 Seventh St., Wright-Patterson Air Force Base, OH 45433), Robert H. Gilkey (Wright State Univ., Dayton, OH 45435), Douglas S. Brungart, Nandini Iyer, and Griffin D. Romigh (Wright-Patterson Air Force Base, OH 45433)

Several studies have shown that detection of a dichotic target signal in a concurrent diotic noise masker improves when the masker onset occurs prior to that of the target and/or the masker duration extends beyond the target offset (so-called masker fringe; McFadden, 1966; Trahiotis *et al.* 1972). It has been argued that this improved performance results from the fact that the masker fringe provides a baseline set of interaural parameters against which a change in parameters, associated with the signal onset or offset, can be detected. As the duration of the fringe increases, more time is available to estimate these baseline parameters, thus improving signal detectability. However, little is known about the relation between masker fringe and spatial hearing. In this study, the localization of a 250-ms click train signal presented in a broadband directional noise masker was measured as a function of the duration of the masker fringe and the signal-to-noise ratio. The results indicate that the presence of masker fringe improves localization performance relative to the case in which the masker and signal are pulsed on and off simultaneously. Moreover, the benefit of masker fringe increases with decreasing signal-to-noise ratio. [Work supported by a grant from AFOSR.]

1pPP2. The effect of flanking bands on sensitivity to interaural time differences. Christopher A. Brown and William A. Yost (Dept. of Speech and Hearing Sci. Arizona State Univ., P.O. Box 870102, Tempe, AZ 85287-0102)

We are examining the potential for cochlear implant patients who retain some low-frequency residual hearing in both ears to benefit from interaural time differences (ITDs) to improve speech reception. Given the likelihood of asymmetrical loss in the low-frequency region, the relationship between spectral overlap and ITD sensitivity is of interest. We previously measured ITD sensitivity using narrow-band noise carriers, while varying the amount of spectral overlap between the ears. We found very good sensitivity when the bands completely overlapped, but for most listeners, sensitivity declined as the center frequency of one band was moved away from the other. When, for example, 2/3-octave bands were offset by 1/3 octave and presented to each ear, ITD sensitivity declined. This condition consists of three bands, a 1/3-octave center band that is spectrally similar across ears, and two 1/3-octave flanking bands, each of which is present in one or the other ear. Because the center band alone produces good ITD sensitivity, we hypothesized that moving the flankers away in frequency from the center band would result in improved sensitivity. However, the presence of the spectrally displaced bands led to poorer ITD sensitivity. [Work supported by grants from NIDCD.]

1pPP3. Does random temporal jitter reduce onset dominance in temporal weighting functions for high-rate click trains? Andrew D. Brown and G. Christopher Stecker (Dept. of Speech & Hearing Sci. , Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105, andrewdb@u.washington.edu)

Several studies have noted poor sensitivity to interaural time differences (ITD) among users of bilateral cochlear implants (CIs) compared to normal hearing listeners. One hypothesis for this impairment is that CI users are subject to a higher degree of binaural adaptation—a loss of sensitivity to binaural cues after onset—than normal hearing listeners, as a result of CI electrical stimulation. [Laback and Majdak, PNAS **15**, 814–817 (2008)] demonstrated improved ITD discrimination in CI users by imposing random temporal jitter on CI pulse trains bilaterally. A candidate explanation for this improvement is that temporal jitter causes a "restarting" of the adapted binaural system, and, thus, improved sensitivity to ITD after onset. In the present study, temporal weighting functions of normal-hearing listeners were measured for discrimination of ITD and interaural level differences carried by trains of 4000 Hz Gabor clicks. Interclick intervals of 1.25 and

2.5 s were tested with and without additional random temporal jitter. The results have implications for mechanisms of binaural adaptation as well as for binaural performance in CI users.

1pPP4. Modeling temporal weighting of interaural time and level differences in high-rate click trains. G. Christopher Stecker and Andrew D. Brown (Dept. of Speech & Hearing Sci. , Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105, cstecker@u.washington.edu)

Accurate localization of sound sources requires that information be integrated over time and across multiple spatial cues in a manner that reflects the reliability of those cues over time. Several past studies have investigated the relative weighting of multiple cues (e.g., interaural time-intensity "trading") as well as the temporal weighting applied to brief stimuli such as noise bursts and impulse trains. The results have demonstrated reduced sensitivity to ongoing envelope interaural time differences in high-frequency, high-rate stimuli and a consequent shift in weighting to favor onset and interaural-level cues. The dominance of onset over ongoing cues is consistent with the precedence effect [Wallach *et al.*, Am. J. Psych. **62**:315–336 (1949)] and binaural adaptation [Hafter and Dye, J. Acoust. Soc. Am. **73**: 644–651 (1983)] and may reflect a combination of (1) peripheral auditory factors, (2) effective modulation, (3) postonset inhibition of binaural information, and (4) temporal integration of ongoing cues. In this study, simple mathematical models are used to study the effects of these factors on temporal weighting of interaural time and level differences in narrowband click trains varying in spectral composition, click timing, and cue availability.

1pPP5. The effectiveness of multiple harmonics in eliciting the precedence effect in normal-hearing listeners. John Ellison and Jonas Braasch (School of Architecture, Rensselaer Polytechnic Inst., Greene Bldg., 110 8th St., Troy, NY 12180, ellisj2@rpi.edu)

Localization dominance due to the precedence effect refers to the ability of listeners to localize a primary auditory event in the presence of reflected sound. It has been shown that localization dominance becomes more stable as the bandwidth of a noise burst increases, which suggests the role of cross-frequency interactions on localization dominance. Currently, little is known about how stimuli containing multiple harmonics influence localization dominance. The present investigation uses a multiple harmonic complex to investigate localization dominance due to the precedence effect. Listeners locate the perceived lateral position of stimuli presented binaurally via headphones by adjusting the interaural-level difference (ILD) that approximates the location of the stimulus. Stimuli are presented as lead and lag signals that simulate direct and reflected sound, respectively. The interstimulus interval (ISI) is varied from 0 to 4 ms, and the phase of the harmonic complex is varied between ears. For both the test stimulus and broadband noise bursts (control condition) physical ILDs occur as a result of lead/lag interference, but in the case of the harmonic complexes, the ILD often does not average out over frequency, as is the case for broadband noise.

1pPP6. Are there specialized mechanisms for pitch interval perception? Josh H. McDermott, Michael V. Keebler, and Andrew J. Oxenham (Dept. of Psych., Univ. of Minnesota, 75 E. River Rd., Minneapolis, MN 55455, oxenham@umn.edu)

Pitch intervals are a central feature of most musical systems. It has seemed likely that pitch uniquely permits precise perception of the interval by which two sounds differ, perhaps due to mechanisms specialized for music. We tested this notion by measuring interval discrimination thresholds for pitch, brightness, and loudness in the same subjects. We compared thresholds across dimensions by computing: (1) Weber fractions (the ratio of the interval threshold to the base interval size) and (2) the ratio of the interval threshold to the basic discrimination threshold for a dimension. We found that Weber fractions were comparable across dimensions (although Weber's law did not hold, as thresholds were roughly constant across the range of interval sizes used), and that the threshold ratio was actually higher (worse) for pitch than for loudness or brightness. Interval thresholds thus provide no evidence that pitch interval perception has unusually high fidelity, as might be expected if it were the product of a specialized mechanism. The ubiquity of pitch in music may simply be due to basic discriminability, which, in contrast to interval discriminability, is far superior for pitch than for other dimensions. [Work supported by NIH grant R01DC05216.]

1pPP7. Exploring the acoustic basis of consonance using individual differences. Josh H. McDermott, Andriana J. Lehr, and Andrew J. Oxenham (Dept. of Psych., Univ. of Minnesota, 75 E. River Rd., Minneapolis, MN 55455, oxenham@umn.edu)

Despite its central role in music, the origins of the preference for consonant over dissonant intervals remains controversial. We applied an individual differences approach to this question. Subjects were first tested for their aversion to pure-tone beats, and to inharmonic spectra of single complex tones. Consistent individual differences were observed in the magnitude of subjects' aversion to beats and inharmonicity, but the two effects were uncorrelated across subjects, suggesting independent factors. Subjects then rated the pleasantness of musical intervals and three-note chords. If particular acoustic factors determine consonance, stronger sensitivity to those factors should predict stronger musical preferences. In each of two cohorts, only the preference for harmonic spectra was consistently correlated with preferences for consonant over dissonant chords and intervals. Although the strength of a subject's consonance preference was also correlated with their musical experience, so was their harmonicity preference, which explained most of the effect of experience on consonance ratings. The results suggest that much of consonance perception is rooted in the basic acoustic property of harmonicity. Harmonicity is evidently predictive of important structures in Western music, and with exposure to music, listeners learn to attend to it. [Work supported by NIH grant R01DC05216.]

1pPP8. Relative contribution of frequency and duration cues to estimates of frequency change in tone sequences and glides. Molly J. Henry and J. Devin McAuley (Dept. of Psych., Bowling Green State Univ., Bowling Green, OH 43403, mcauley@bgsu.edu)

This research compared use of frequency and duration cues in magnitude estimates of frequency change (velocity). In two experiments, listeners rated how quickly tone sequences (Experiment 1) or glides (Experiment 2) changed in frequency on a scale ranging from 0 (not changing at all) to 100 (changing very quickly). In both experiments, velocity was specified in terms of semitones per second (ST/s), with ascending and descending stimuli presented at one of eight velocities on each trial (0, 2, 4, 6, 8, 10, 12, and 14 ST/s) with a variable starting frequency. Frequency and duration cues to velocity were varied orthogonally. The different stimulus velocities ($\Delta f / \Delta t$) were created by either holding Δt constant and varying Δf (frequency cue condition) or by holding Δf constant and varying Δt (duration cue condition). For tone sequences, magnitude estimates of velocity were based almost entirely on duration cues, with very little contribution of frequency cues. This finding contrasted with magnitude estimates of glides where both duration and frequency cues were equally effective at conveying velocity information.

1pPP9. Gender differences in the ability to ignore distracters in frequency discrimination tasks. Blas Espinoza-Varas (Dept. Commun. Sci. & Disord., Univ. Oklahoma Health Sci. Ctr., Oklahoma City, OK 73117–1215) and Praveen Jajoria (Univ. of Texas Medical Branch, Galveston, TX 77555-0570)

Whether there are gender differences in the ability to ignore distracting auditory information remains a matter of debate; such differences could be in basic, involuntary, novelty responses or in elaborate, voluntary, directed attention. To investigate this issue, frequency discrimination thresholds (FDTs) were measured in 20 males and 20 females (young, healthy adults with normal hearing). Targets tones (T) were presented in isolation or followed by a distracter tone (D) after a 90- or a 250-ms silent interval (ISI). Both tones were 1000 Hz, 70 dB SPL, and duration was 40 and 80 ms for T and D, respectively. Target frequency differences were specified by an adaptive rule targeting 71 percent-correct FDTs; irrelevant differences in D frequency were 40 or 45 Hz. On each subject, FDTs were measured for T in isolation or followed by D, with or without irrelevant differences; a total of 12 trial blocks per condition were completed. Across gender, the largest FDT disparity was observed in conditions with irrelevant differences in D

frequency and 90-ms ISI; in the other conditions the gender effect was small or negligible. There appears to be a gender difference in the initial, involuntary, response to distracters.

1pPP10. The relative weighting of energy and envelope-modulation cues in increment detection. Harisadhan Patra, Jessica J. Messersmith, and Walt Jesteadt (Boys Town Natl. Res. Hospital, 555 N 30th St., Omaha, NE 68131, patrah@boystown.org)

Recent findings suggest that listeners primarily rely on envelope-modulation cues rather than energy cues in an increment detection task, where one of the two sounds contains a brief increment in intensity in a longer duration tone. The relative contributions of energy and envelope-modulation cues in increment detection were evaluated in six normal-hearing young adults. The detection thresholds for a 20 ms increment added in the middle of a 420 ms pedestal of either 500 or 4000 Hz were obtained as a function of pedestal level. On-frequency or off-frequency maskers were either 30- or 150-Hz wide random-phase Gaussian noise (RPN) or low-noise noise (LNN), centered either at 500 or 4000 Hz. The envelope fluctuation in LNN is minimal compared to that in RPN. The detection thresholds were greater in RPN than LNN in on-frequency, but not in off-frequency masking conditions. The detection thresholds increased at the wider masker bandwidth. Results suggest that listeners may rely on a decision process that is based on within-channel energy and envelope-modulation cues, where the within-channel envelope-modulation cues predominate. The relative weighting of these cues are pedestal and masker dependent. [Work supported by R01 DC006648 and T32 DC000013.]

1pPP11. Effect of temporal uncertainty on psychometric functions for increment detection. Walt Jesteadt, Harisadhan Patra, and Donna L. Neff (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

Increased temporal uncertainty might be expected to produce increased internal noise, resulting in poorer detection and shallower psychometric functions. An alternative model in the literature predicts steeper psychometric functions with increased uncertainty. As a test of these two alternatives, detection of a 20-ms intensity increment was assessed for eight subjects in a two-interval, forced-choice adaptive procedure where the increment was presented 100, 150, 200, 250, or 300 ms after the onset of a 400-ms, 4-kHz tone presented at 70 dB SPL. In one condition, the temporal position varied randomly from trial to trial. In the other, the temporal position was fixed for an entire 50-trial block. Detection thresholds were higher and psychometric functions obtained by analyzing the trial-by-trial data from the adaptive tracks were shallower in the higher temporal uncertainty condition. Thresholds in both conditions were higher for temporal positions more distant from the center of the 400-ms tone. Results support a model in which temporal uncertainty contributes to internal noise resulting in shallower psychometric functions. [Work supported by R01 DC006648 and T32 DC000013.]

1pPP12. Increment discrimination. Walt Jesteadt, Harisadhan Patra, and Jessica Messersmith (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

Subjects rely on different information in an intensity-discrimination task, where the two sounds to be compared differ only in intensity, than in increment detection, where one of the two sounds contains a brief increment in intensity in a longer duration tone. To explore the transition from one task to the other, ten subjects were tested in a two-interval forced choice discrimination task in which both intervals contained 50-ms increments in a 400-ms, 500-Hz tone and subjects were asked to vote for the interval with the larger increment. The size of this second increment was adapted to obtain an estimate of threshold. Increment-discrimination thresholds were better than increment-detection thresholds when the standard increment was less than or equal to detection threshold. Increment-discrimination thresholds became poorer for standard increments in the range from 1.5 to 12 dB above the level of the longer duration tone. For standard increments above that level, increment discrimination approximated intensity discrimination. Slopes of psychometric functions became shallower as the standard increment increased, consistent with a transition from an increment-detection to an intensity-discrimination decision process. [Work supported by R01 DC006648 and T32 DC000013.]

1pPP13. Effect of narrowband noise maskers on increment detection. Jessica J Messersmith, Harisadhan Patra, and Walt Jesteadt (Hearing Res., Boys Town Natl. Res. Hospital, 555 N 30th, Omaha, NE 68131, messersmithj@boystown.org)

The ability to detect a brief increment in intensity in a longer duration stimulus is often used as a measure of the intensity resolution of the auditory system, but it is unclear how listeners perform the task. Early investigators suggested that detection of an increment in intensity was based on the overall energy within the critical band centered on the signal frequency. The use of signal energy falling outside of the critical band is often limited by introducing a noise masker. Recent findings indicate that envelope fluctuations in a masker can increase thresholds in increment detection, suggesting a decision process based on envelope modulation. A modulation-based decision device was evaluated by adding irregularity to the envelope of an increment detection task via noise maskers of varying bandwidth. Results indicate that thresholds for detection of an increment in the intensity of a pure tone increase with increasing bandwidth for an on-frequency noise masker, but are unchanged by an off-frequency noise masker. A model that includes a modulation filter bank analysis of envelope modulation can account for observed data, but a model based on discrimination of spectral patterns does equally well. [Work supported by R01 DC006648 and T32 DC000013.]

1pPP14. Modulation detection of sawtooth-wave modulated tones: Effect of temporal asymmetry. Yi Shen and Jennifer Lentz (Dept. of Speech & Hearing Sci. , Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405, shen2@indiana.edu)

Large perceptual differences can be observed between ramped and damped sounds. In the present study, the effect of temporal asymmetry on amplitude modulation detection was studied with the use of two types of modulators: a ramped sawtooth wave, characterized by a gradual linear rise in amplitude followed by a sharp drop within each period, and the damped sawtooth wave which is the time reverse of the ramped wave. Results showed that a puretone carrier modulated by these two modulators led to differences in modulation detection thresholds with the damped modulation pattern being easier to detect than the ramped modulation pattern. The size of this effect was also dependent on the modulation rate. A model of auditory processing was proposed to account for the temporal asymmetry in modulation detection with the hypothesis that the asymmetric threshold originated at processing stages more central to the auditory nerve.

1pPP15. Monaural envelope correlation perception and broadband temporal analysis. Joseph W. Hall, III, Emily Buss, and John H. Grose (Dept. of Otolaryngol., UNC Chapel Hill, CB 7070, Chapel Hill, NC 27599-7070, jwh@med.unc.edu)

Monaural envelope correlation perception [V. M. Richards, J. Acoust. Soc. Am. **82**, 1621–1630 (1987)] is a paradigm that reflects sensitivity to the correlation between the temporal envelopes of carriers that are separated in frequency. Previous experiments investigating this paradigm have employed narrowband noise stimuli, presumably because of a desire to ensure that the temporal envelopes of the bands are faithfully represented at the outputs of frequency selective auditory filters. Although wide comodulated bands can be generated, a reasonable expectation is that the representations of the bands at the outputs of auditory filters corresponding to their center frequencies would have substantially reduced comodulation, partly due to the fact that the absolute width of the auditory filter increases with increasing stimulus frequency. Our preliminary data nevertheless indicate robust sensitivity to envelope correlation for noise bandwidths that greatly exceed the bandwidths of peripheral filters. This sensitivity persists even when manipulations are employed that should interfere with the ability to use spectral profile cues or cues based upon envelope comparisons across subsets of

auditory filters. Results will be discussed in terms of cues supporting envelope correlation perception with stimuli that are broad with respect to auditory filters. [Work supported by NIDCD 5R01DC000418.]

1pPP16. Randomized phase effects in amplitude modulated/ quasi-frequency modulated discrimination. Hisaaki Tabuchi, Ewa Borucki, and Bruce G. Berg (Dept. of Cognit. Sci., Univ. of California, Irvine, 3151 Social Sci. Plaza, Irvine, CA 92617-5100)

This study investigated the bandwidth of phase sensitivity (i.e., discrimination of waveform envelopes). Subjects discriminated amplitude modulated (AM) tones, and quasi-frequency modulated (QFM) tones in a 2IFC. An adaptive threshold procedure was used to estimate modulation depth m needed to make the discrimination as a function of carrier and modulation frequency. Nonmonotonic threshold functions were often observed at higher modulation frequencies. The effect is discussed in terms of two potential cues: (1) waveform envelope, (2) cubic distortion products. In order to degrade the information obtained from the distortion products, the phase for the carrier frequency was randomly sampled with a range of, $\Delta\theta = \pi/3$. The randomized phase diminishes the nonmonotonicities of the threshold functions. Final results show that bandwidths of maximum sensitivity are not proportional to carrier frequencies. There are also noticeable individual differences in both the effects of the distortion products and the maximum bandwidths of sensitivity. [Work supported by NSF BCS-07464003.]

1pPP17. The range of signal-to-noise ratios needed to equalize identification of different consonants. E. William Yund, Pierre L. Divenyi, Matthew A. I. Ua Cruadhlaoich (Res. Service, VA NCHCS, 150 Muir Rd., Martinez, CA 94553, yund@ebire.org), and David L. Woods (Neurology Dept., VA NCHCS & UC Davis, Martinez, CA 94553)

Individual consonants vary in intelligibility when speech is presented in constant-intensity speech-spectrum noise. This variation in consonant intelligibility is explained at various levels, ranging from bottom-up energetic masking of different-intensity auditory cues needed to identify different consonants, to the top-down use of context information. This study focuses on bottom-up auditory cues by measuring the differences in signal-to-noise ratio (SNR) needed to equate consonant identification in consonant-vowel-consonant (CVC) syllables. The goal is to develop a CVC test with high sensitivity to changes in consonant perception because each consonant is presented at SNRs above the perceptual floor and below the perceptual ceiling. In order to present the initial and final consonants of each CVC at independently controlled SNRs, the intensity of the noise was changed during 100 ms centered in the vowel. In 16 young normal-hearing listeners, the SNR range required to equate identification across the 20 leading and trailing consonants was greater than 40 dB. In spite of these large SNR variations occurring during the test (and sometimes even within a single CVC) the expected pattern of consonant confusions was obtained, suggesting that the mid-vowel adjustment of noise intensity did not interfere with normal consonant-perception processes.

1pPP18. Relationships between timing in reading and language ability and auditory temporal masking. Cynthia M. Zettler (Waisman Ctr., Univ. of Wisconsin-Madison, 1500 Highland Ave., Madison, WI 53705), Rose A. Sevcik, Robin D. Morris, Mary Ann Romski, Colleen O'Rourke, and Marsha G. Clarkson (Georgia State Univ., Atlanta, GA 30303)

Broad timing difficulties may play a role in reading problems, but its role and a potential relationship with auditory temporal processing are unclear. The present study explored relationships between variables that might identify timing as a possible link between reading and auditory temporal masking. Tone identification thresholds were measured for 75 adults for backward masking and comodulation masking release (CMR) stimuli. Adults also completed a battery of nonverbal IQ, naming speed, motor speed, language, and reading measures. Backward masking stimuli consisted of a 20-ms, 1-kHz tone immediately preceding a 300-ms, 600–1400 Hz bandpass noise masker. A notched-noise condition included a spectral notch between 800 and 1200 Hz. CMR stimuli consisted of a reference condition in which a 1-kHz tone was temporally centered in a 20-Hz wide narrowband modulated noise masker alone and one with eight flanking bands of modulated noise in adjacent frequency regions present. CMR was measured as the improvement in threshold with flanking bands present relative to the single-masker reference. Auditory stimuli were presented via headphones. Results indicated that nonverbal intelligence predicted performance on masking tasks, supporting previous research. Likewise, motor speed predicted performance for some of the masking tasks, suggesting that the underlying timing elements are not independent.

1pPP19. Is intelligibility of adjacent passbands hypoadditive or hyperadditive? Richard M. Warren, James A. Bashford, Jr., and Peter W. Lenz (Dept. of Psych., Univ. of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201)

Based on their own findings and reports from other laboratories, H. Müsch and S. Buus [H. Müsch and S. Buus, J. Acoust. Soc. Am. **109**, 2896–2909 (2001)] suggested that when heard together, the intelligibilities of adjacent passbands were hypoadditive, and those of disjoint passbands were hyperadditive. A subsequent study employed extremely high order far infrared filtering that had been shown to effectively eliminate contributions from transition band slopes [R. M. Warren, J. A. Bashford, Jr., and P. W. Lenz, J. Acoust. Soc. Am. **118**, 3261–3266 (2005)]. That study measured the intelligibility for each of the 15 possible pairings of six one-octave effectively rectangular passbands (3 dB/Hz filter skirts) that spanned the speech spectrum with center frequencies ranging from 0.25 to 8 kHz. Each pairing, whether contiguous or disjoint, exhibited hyperadditivity. The present study determined whether decreasing the filter skirts to 0.5 dB/Hz (considered quite steep by conventional standards) would produce the hypoadditivity reported in literature for adjacent bands. Results obtained support the hypothesis that redundancy introduced by overlapping transition band slopes could be responsible for the redundancy correction factor employed by some models for estimating intelligibility of paired adjacent passbands. [Work supported by NIH.]

1pPP20. Auditory stream segregation in six-month-old infants and adults. Nicholas A. Smith (Boys Town Natl. Res. Hosp., 555 North 30th St., Omaha, NE 68144, smithn@boystown.org) and Laurel J. Trainor (McMaster Univ., Hamilton, ON L8S 4K1, Canada)

Auditory stream segregation in six-month-old infants was examined in a paradigm adapted from Bregman and Rudnicky (1975). In a conditioned head-turn procedure, listeners' detection of changes in the temporal order of 2200 and 2400 Hz target tones was tested under three conditions. For developmental comparison, as well as validation of the stimulus design, adults were tested in an identical procedure. Adults clearly detected changes in the order of the target tones when presented alone, but performance suffered when the target tones were preceded and followed by 1460 Hz flanker tones. However, performance improved when the flanker tones were embedded within a series of additional 1460 Hz tones, which captured the flanker tones into an isochronous stream of tones, perceptually segregated from the targets. Although infants performed worse than adults under the same conditions, infants also benefited from the addition of captor tones, suggesting that they too can use grouping cues to organize and attend to elements in the auditory scene.

1pPP21. Effect of selective attention on the segregation of interleaved melodies. Aymeric Devergie, Nicolas Grimault, Barbara Tillmann (UMR CNRS 5020, Univ. Lyon 1, 50 av T Garnier 69366 Lyon Cedex 09 France), and Frédéric Berthommier (GIPSA-LAB, INPG, Grenoble, France)

The role of attention in streaming remains controversial and the data from the literature appear at first sight to be inconsistent [C. Alain and L. J. Bernstein, Curr. Opin. Otolarygol. Head Neck Surg. **16**, 485–489 (2008)], aimed to reconcile these data arguing that selective attention may increase segregation via a sensory mechanism that enhances the processing of task-relevant acoustic information, and a suppression mechanism that attenuates the processing of task-irrelevant stimuli. To test this hypothesis, we proposed an interleaved melody paradigm with various rhythm conditions applied to familiar target melodies and unfamiliar masking melodies. All targets and maskers had the same pitch ranges and timbre variations. No rhythm cue was available in condition 1 (irregular target and masker rhythms). The masker was presented at a regular rhythm in condition 2, the target remaining irregular. Performance over chance in condition 1 indicated that attention enables listeners to extract a familiar melody from a background melody. The better performance in condition 2 can be interpreted in terms of suppression of the masker melody based on the rhythm regularity.

Overall, the current findings are consistent with Alain and Berstein's hypothesis that selective attention can modulate neural processes involved in auditory scene analysis.

1pPP22. Effects of on- and off-frequency maskers on speech recognition. Frederic Apoux and Eric W. Healy (Dept. of Speech and Hearing Sci. The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210, apoux.1@osu.edu)

The present study investigated the relative influence of on- and off-frequency noise bands on speech recognition. The target speech stimuli consisted of 64 consonant-vowel consonants. The background noise was simplified speech spectrum-shaped noise or a sentence played backward. All stimuli were divided into 30 contiguous auditory filter width bands spanning 80 to 7563 Hz. Normal-hearing listeners were presented with 15 target speech bands having frequency locations determined randomly from trial to trial to simulate the quasi-random masking effect of noise on the distribution of speech cues across frequency. Results showed that high levels of noise had little impact on recognition when that noise resided only in the 15 off-target bands. Performance was similar when noise was present in all bands versus when present only in the 15 target bands, further indicating a modest effect of off-frequency noise on speech recognition. Performance with sine-wave vocoded speech bands was systematically lower than that obtained with the unprocessed speech bands, supporting the assumption that fine structure information plays a significant role in speech recognition in noise.

1pPP23. Intelligibility of sequentially presented speech signals. Nandini Iyer, Douglas S. Brungart, and Brian D. Simpson (Air Force Res. Lab., 2610 Seventh St., Bldg. 441, Area B, WPAFB, OH 45433)

In some operational environments, listeners are required to monitor several channels of simultaneous time-critical communication. There are two possible ways to reduce the transmission time for multiple speech messages: (1) present information from two or more speech channels simultaneously or (2) time compress the competing speech messages and present them sequentially. Previous studies have shown that the intelligibility of a single speech message is fairly impervious to high degrees of time compression. However, it is not clear if intelligibility is preserved when several such highly compressed messages are presented sequentially. The current experiment investigated speech intelligibility when two, three, or four time-compressed speech signals were presented sequentially. Three results were apparent: (1) speech intelligibility was near ceiling when time-compressed signals were presented sequentially and listeners were required to identify and respond to only one of the signals in the sequence; (2) speech intelligibility decreased dramatically when the listeners were required to respond to all the messages in the sequence, particularly at high compression rates; and (3) overall performance in the multiple-message identification task depended primarily on the total time interval between the onset of each new message, and not on the degree of time compression of the speech message contained in that interval.

1pPP24. Further studies on the benefit of a carrier phrase for word recognition in a two-talker masker: Contributions of coarticulation and auditory stream segregation. Angela Y. Bonino, Lori J. Leibold (Dept. of Allied Health Sci., Univ. of North Carolina, Campus Box 7190, Chapel Hill, NC 27599, abonino@med.unc.edu), and Emily Buss (Univ. of North Carolina, Chapel Hill, NC, 27599)

In adults with normal hearing, word recognition has been shown to benefit from inclusion of a carrier phrase in the presence of a continuous two-talker masker, but not with speech-shaped noise [Bonino *et al.*, Association for Research in Otolaryngology MidWinter Meeting, Baltimore, MD (2009)]. One explanation for this finding is that a carrier phrase aids in the perceptual segregation of the target word from the two-talker background, either through auditory stream segregation or through coarticulation of the carrier and target. Two experiments were conducted to investigate the benefit of the carrier phrase, "say the word," presented prior to a monosyllabic target word. In experiment 1, performance was assessed as a function of signal-to-noise ratio (SNR) in either a two-talker masker or a speech-shaped noise, presented at 0, +5, and +10 dB SNR. In experiment 2, performance was compared across two carrier conditions utilizing either the carrier/target pairs from the original recording or a single utterance of the carrier preceding all target presentations; the time delay between the carrier phrase and target word was also manipulated. Results from these experiments will be discussed, with a focus on the implications for understanding the underlying mechanism responsible for the carrier phrase benefit. [This work was supported by NIH NIDCD grant R03 DC008389 and American Academy of Audiology/American Academy of Audiology Foundation Research.]

1pPP25. Effects of training format on perceptual learning of spectrally degraded voices. Sangsook Choi, Karen Kirk (Dept. of Speech, Lang., and Hearing Sci., Purdue Univ., 500 Oval Dr., West Lafayette, IN 47907, choi111@purdue.edu), Tom Talavage Vidya Krull Chris Smalt, and Steve Baker (Purdue Univ., West Lafayette, IN 47907)

There is evidence that familiarity with the talker facilitates speech recognition (see Nygaard, 2005 for reviews). Identifying talker voices from an auditory speech signal is difficult for cochlear implant users due to the lack of spectral detail conveyed by their processors. Recently, Sheffert and Olson (2004) demonstrated facilitation in voice learning when listeners with normal hearing were trained using audiovisual speech. In the present study, normal hearing listeners were presented with speech processed through a 12-channel noise band cochlear implant simulator and trained to recognize voices in an auditory-only or audiovisual format. Listeners in the audiovisual format required fewer sessions to reach criterion (80% accuracy) than those in the auditory-only format. Voice learning varied as a function of the talkers included in the training set. For the talkers that were more difficult to identify, audiovisual training yielded greater improvements in voice identification than did auditory-only training. After training, participants completed a word recognition task using processed sentences produced by training talkers and novel talkers. Talker familiarity did not influence word recognition performance for either training group. This may be due to ceiling effects with the 12-channel noise band processor. Additional testing in noise is underway.

1pPP26. Spectral modulation detection training: Different outcomes for different trained stimuli. Andrew T. Sabin (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr. Evanston, IL 60201, a-sabin@northwestern.edu), David A. Eddins (Univ. of Rochester, Rochester, NY 14618), and Beverly A. Wright (Northwestern Univ., Evanston, IL 60201)

The pattern of peaks and valleys of sound level spread across audio frequency (spectral modulation) is a key stimulus cue used for sound identification and localization. However, little is known about how practice can influence the ability to detect this modulation. Here we trained three separate groups of normal-hearing adults to detect one of three different spectral modulations. Listeners practiced distinguishing a noise with a flat spectrum from one with a sinusoidal spectral modulation 1 hr/day for 7 days. We varied the modulation depth adaptively to determine the detection threshold. On average, the groups trained to detect 0.5 cyc/oct or 1 cyc/oct spectral modulation across a 200–1600 Hz carrier improved significantly, while the group trained to detect 2 cyc/oct spectral modulation across a 400–3200 Hz carrier did not. Further, it took multiple sessions for the 0.5-cyc/oct group to reach optimal performance, but only one session for the 1-cyc/oct group. Thus it appears that the current training had a smaller and more rapid effect as spectral modulation frequency increased from 0.5 to 2 cyc/oct (if we assume no influence of carrier). Pre-training thresholds also decreased over this range, suggesting that there is less room to improve at higher spectral modulation frequencies.

1pPP27. Difference limen for dynamic center-of-gravity signals. H. Dean Hudson (Air Force Res. Lab, 2610 Seventh St., Bldg 441, Area B, WPAFB, OH 45431), Evelyn M. Hoglund, and Lawrence L. Feth (Columbus, OH 43210)

We hypothesize that central auditory processing of the dynamic center-of-gravity (COG) of a pair of tones is similar to that for a single tone that undergoes sinusoidal frequency modulation (FM). The signal COG can be moved dynamically by introducing amplitude modulation to two fixed-frequency anchor tones separated by one, or more, equivalent rectangular bandwidths (ERB). The phase relationship between the sinusoidal amplitude modulators of the anchor tones is used to control the magnitude of the COG deviation. In this study, the amplitude of the anchor tones and the depth of AM were identical between intervals. The phase relationship of the modulators was manipulated to control the deviation in the COG. To test our hy-

pothesis, we compared FMDLs with COGDLs at three center frequencies and three modulation rates. First, difference limens were measured for the FM signals in partial replication of [Ozimek and Sek (1990)]. Next, COGDLs were measured. Initial attempts to match COG deviation to FM deviation were unsuccessful because the task was too difficult. Consequently, the modulation depth of the COGDL signals was fixed at 0.5, which resulted in replicable performance across subjects. Data for both the FMDL and COGDL task will be discussed. [Work supported by NIH and ASEE.]

1pPP28. Distortions in the perceived relative temporal positions of a signal and masker. Beverly A. Wright (Dept. of Comm. Sci. and Disord., Northwestern Univ., 2240 Campus Dr. Evanston, IL 60208, b-wright@northwestern.edu) and Matthew B. Fitzgerald (New York Univ., New York, NY 10016)

Simultaneous and non-simultaneous masking tasks are often introduced to listeners in terms of the veridical relative positions of the signal and masker. Here we report that the actual perceived temporal positions of the signal and masker do not always match those descriptions. We asked normal-hearing adults about the perceived presentation time of a brief "signal" that was presented either before, during, or after a longer "masker." The task was either to mark the perceived presentation time of the signal on a schematic diagram of the masker, or simply to judge whether the signal occurred inside or outside of the masker. Listeners tended to perceive the signals as having occurred later than they actually had. This temporal distortion was most marked for signals that were presented near the end of the masker. Signals that were presented 25–100 ms before the end of the masker were perceived as occurring outside of the masker 50% of the time. Similar results were obtained with different masker durations (150, 300, and 450 ms), and various combinations of signal and masker spectra. This phenomenon may provide a measure of the time it takes to switch attention from a foreground to a background stimulus. [Work supported by NIH.]

1pPP29. Detection of warning signals while operating a motorcycle and wearing hearing protection. Andrew D. DeSherlia (6643 N. Newgard, Chicago, IL 60626, andyds23@gmail.com)

It has been shown that operators of motorcycles are exposed to dangerous levels of wind noise at high speeds [McCombe *et al.*, J. Laryngol. Otol. **109**, 599–604 (1995)]. Many governmental regulatory bodies have developed restrictions that ban the use of earplugs while operating a motorcycle, due to their obstruction of warning signals. In order to test this theory, 20 people with normal hearing were exposed to recordings of motorcycle wind noise accompanied by warning signals at different amplitudes. As the warning signal's amplitude was adjusted, the listener was asked to indicate when it could be detected. This determined the threshold at which the warning signal could be detected for listeners with and without hearing protection. This study should broadly indicate whether hearing protection allows the listener to detect warning signals at a lower threshold than if no protection were used.

1pPP30. Model-based direction estimation of concurrent speakers in the horizontal plane. Mathias Dietz, Stephan D. Ewert, and Volker Hohmann (Medizinische Physik, Univ. Oldenburg, Germany)

A computational auditory model for the direction estimation of concurrent speakers in the horizontal plane is presented. The auditory signal processing is motivated by mammalian physiology and human psychoacoustic data, extending the binaural model by Dietz *et al.* [Brain Res., **1220**, 234–245 (2008)]. The model exploits a wide range of interaural parameters, namely, the interaural level difference, the interaural phase difference of the fine structure, and the interaural coherence. The parameters are determined at the output of auditory filters allowing for a simultaneous estimation of several sound source directions in a multidimensional internal representation. By focusing on time segments with high interaural coherence [Faller and Merimaa, J. Acoust. Soc. Am. **116**, 3075–3089 (2004)] in the auditory bands, which likely stem from a localized sound source, the robustness of the direction estimation of the model was considerably improved: for two well-separated speakers in quiet, the estimation error is generally less than or equal to 10 deg. In noisy environments, the direction of a single speaker can be estimated for signal-to-noise ratios below 0 dB. Further improvements of the model by concurrent pitch estimation to reduce permutation errors between nearby speakers are investigated and discussed.

1pPP31. A comparison of modulation indices for describing amplitude-compressed speech. Pamela Souza (Dept. of Speech and Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105, psouza@u.washington.edu), Frederick Gallun (Portland VA Medical Ctr., Portland, OR 97239), and Eric Hoover (Univ. of Washington, Seattle, WA 98105)

This presentation explores two approaches to characterizing changes to speech with amplitude-compression amplification. In the first, we considered changes to waveform morphology using the envelope difference index (EDI); [Fortune *et al.* (1994)] to quantify the effects on amplitude envelope in the time domain. In the second, the spectral correlation index (SCI); [Gallun & Souza (2008)] was used to compare signal envelopes in the frequency domain. In the SCI, similarity is represented by the correlation of the average modulation spectra at six octave-spaced amplitude modulation frequencies (1–32 Hz) in six octave-spaced carrier frequency regions (250–8000 Hz) for a pair of phonemes. Test stimuli were a set of 16 vowel-consonant-vowel (VCV) syllables spoken by four talkers. The VCVs were digitally processed to limit spectral information using a signal-correlated noise technique [Schroeder (1968)] then submitted to a compression amplification simulation in which parameters (compression ratio and release time) were selected to achieve a range of temporal alteration. Consonant identification scores for 10 normal-hearing listeners were obtained and related to the modulation indices. Although the SCI was more sensitive to changes in specific compression parameters, the EDI was more closely related to recognition. [Work supported by R01 DC0060014 and VA RR&D CDA-2 C4963W.]

1pPP32. Extending the sawtooth wave form inspired pitch estimator with an auditory-based preprocessing stage. Arturo Camacho (Escuela de Ingen. Eléct., Univ. de Costa Rica, Apdo. Postal 11501-2060 UCR, San José, Costa Rica)

The sawtooth wave form inspired pitch estimator (SWIPE) [A. Camacho, J. Acoust. Soc. Am. **124**, 1638–1652 (2008)] is extended by introducing an auditory-based preprocessing stage consisting of two main steps. First, a system consisting of a gammatone filterbank followed by half-wave rectifiers and low-pass filters is applied to the signal. This system tends to boost low-order harmonics of the signal, as shown in a work by Klapuri [A. Klapuri, IEEE Trans. Audio, Speech, Lang. Process. **16**, 255–266 (2008)]. Second, the square root of the amplitude spectrum is computed on every channel, and they are summed to form a *summary square-root spectrum* (SSRS). SWIPE is then applied to this SSRS. Results show that the proposed method significantly outperforms the original SWIPE on music and telephonelike speech, albeit producing a small decrease in performance on high-fidelity speech.

1pPP33. Head shadow effect and pinna effect by sound intensity. Yen-Sheng Chen (Taiouan Interdisciplinary Otolaryngol. Lab., Chang Gung Univ.), Jen-Fang Yu (Chang Gung Univ.,), Chin-Kuo Chen (Chang Gung Memorial Hosp.), Wei-De Cheng, Yi-Hsin Lin (Chang Gung Univ.), and Go-Long Tsai (Natl. Taipei Univ. of Technol., Taipei, Taiwan)

This study is to develop a procedure of measuring the influences of the head shadow effect and pinna effect by stimulus intensities and stimulus frequencies (125 Hz to 8 kHz) based on the stimulus source at azimuth angles (0°–180°). In this study, the head shadow effect would be affected by the stimulus frequencies at 3–8 kHz, and the difference of sound pressure for bilateral ears was 7.46–41.40 dB SPL. The pinna effect would be affected by the stimulus frequencies at 3–8 kHz. The difference of sound pressure for all sound sources was 6.62–35.69 dB SPL. The pinna effect would occur most apparently at 135° of azimuth. Consequently, the sound intensity would not affect the head shadow effect and the pinna effect. On the con-

trary, the pinna effect, the head shadow effect, the frequency of stimuli and the sound location should be considered for hearing aids selection for hearing-impaired listeners.

1pPP34. Matching the waveform and the temporal window in stimulus generation. William M. Hartmann (Dept. of Phys. and Astron., Michigan State Univ., 1226 BPS Bldg., E. Lansing, MI 48824 hartmann@ps.msu.edu) and H. Steven Colburn (Boston Univ., Boston, MA 02215)

The discrete Fourier transform (DFT) of a signal that is windowed with a rectangular window having duration T_D is given by the coefficients of the Fourier series for a periodic waveform having period T_D. This periodic waveform is "matched" to the rectangular window because of a special relationship between the components of the windowed signal (windowed sine or cosine functions) and the harmonics of the periodic waveform (sine or cosine functions). Component m of the windowed signal is orthogonal to harmonic n of the waveform ($m \neq n$). If the window (overall duration T_D) is not rectangular, the DFT still corresponds to a periodic waveform having period T_D, but that waveform is not matched to the window because the components of the windowed signal are not orthogonal to the harmonics of the periodic waveform. If the window has raised-cosine edges, with durations τ, on either end, the components of the windowed signal are orthogonal to harmonics of a periodic waveform only if the period of the waveform is $T_D - \tau$, not T_D. The distinction becomes important in generating stimuli for psychoacoustical and physiological experiments. [Work supported by the NIDCD, Grants DC00181 and DC00100.]

1pPP35. An efficient code for environmental sound classification. Robert Lutfi (Dept. of Communicative Disord., Univ. of Wisconsin, Madison, WI 53706, ralutfi@wisc.edu), Raman Arora, and Ching-Ju Liu (Univ. of Wisconsin, Madison, WI 53706, chingjuliu@wisc.edu)

Much work has been devoted in recent years to the goal of developing an automated sound recognition system that can accurately and efficiently classify a wide variety of common environmental sounds according to their generating source. The effort is driven, in part, by the desire to understand our own ability to reconstruct from sound an accurate perception of everyday objects and events present in our natural environment [cf. A. S. Bregman, *Auditory Scene Analysis*, MIT Press (1990)]. Here an efficient code is described for classifying environmental sounds that exploits a recent significant advance in sparse signal reconstruction known as compressed sensing (CS) [cf. D. L. Donoho, IEEE Trans. Inf. Theory, **52**, 1289–1306 (2006)]. CS involves a novel approach to sampling in which the salient information in signals is recovered from the projection onto a small set of random basis functions. The advantage of CS over traditional Fourier or wavelet representations is its capacity to allow accurate classification at low signal-to-noise ratios based on few samples and little or no prior information about signals. To demonstrate we compare CS-based and Fourier-based classification for several representative environmental sound classification tasks. [Work supported by NIDCD].

1pPP36. A computational model for auditory saliency of environmental sound. Bert De Coensel (Dept. of Elec. Eng. and Comput. Sci., Univ. of California at Berkeley, 415 Soda Hall, Berkeley, CA 94720-1776, bert.decoensel@intec.ugent.be), Dick Botteldooren (Ghent Univ., B-9000 Ghent, Belgium), Birgitta Berglund, and Mats E. Nilsson (Stockholm Univ. and Karolinska Inst., SE-106 91 Stockholm, Sweden)

Because the information flow received by the human auditory system exceeds the processing capacity of the brain, neural mechanisms engage and guide attention toward prominent parts of the auditory scene. Several computational models for auditory saliency have been proposed recently. Most of these are concerned with speech recognition, and therefore apply high temporal and spectral precision to relatively short sound fragments. Here, a simplified model is described that specifically targets the long exposure times usually considered in soundscape research. The model trades temporal and spectral accuracy for computational speed, but nevertheless implements the key elements that are present in the calculation of complex auditory saliency maps. A simplified "cochleagram" is calculated from the 1/3-octave band spectrogram using the Zwicker model for specific loudness. Saliency is determined based on spectro-temporal irregularities, extracted in parallel at different feature scales, using a center-surround mechanism. Finally, conspicuous peaks are selected using within-feature and between-feature competitions. The model is shown to behave as expected for a number of typical sounds. As an illustration, saliency calculation results for a set of recordings in urban parks are compared with other acoustical descriptors and with perceptual attribute scales from questionnaire studies.

1pPP37. Signal-processing strategies for electric-acoustic stimulation. Christopher A. Brown and Sid P. Bacon (Dept. of Speech and Hearing Sci. Arizona State Univ., PO Box 870102, Tempe, AZ 85287-0102)

Electric-acoustic stimulation (EAS) occurs when a cochlear implant recipient receives some amount of acoustic stimulation, typically as a consequence of residual low-frequency hearing. This combined stimulation often leads to significant improvements in speech intelligibility, especially in the presence of a competing background. We have demonstrated EAS benefit when target speech in the low-frequency region is replaced with a tone modulated in frequency with the target talker's F0, and in amplitude with the amplitude envelope of the target speech. We have hypothesized that in some instances, the modulated tone may be more beneficial than speech. For example, a patient possessing extremely limited residual hearing (e.g., only up to 100 Hz) may not show EAS benefit because the F0 of many talkers is too high in frequency to be audible. In support of this hypothesis, we demonstrated that the tone can be shifted down considerably in frequency without affecting its contribution to speech intelligibility. In the current experiment, while generating the tone, we manipulated various signal-processing parameters to examine their effects on the benefits of EAS. One goal of this work is to develop a real-time processor that provides EAS benefit to implant users with limited residual hearing. [Work supported by NIDCD.]

Session 1pSC

Speech Communication: Clinical, Methodological, and Technological Issues in Speech (Poster Session)

Michelle R. Molis, Chair
National Ctr. for Rehabilitative Auditory Research, Portland, OR 97207

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 even-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

1pSC1. Developing objective measures of listening effort. Mini N. Shrivastav, Rahul Shrivastav, and James D. Harnsberger (Dept. of Commun. Sci. and Disord., Univ. of Florida, 336 Dauer Hall, Gainesville, FL 32611, mnshriv@ufl.edu)

The goal of the present work is to develop valid measures of listening effort independent of task difficulty. Various markers of effort will be measured for listening conditions of varying difficulty but with equal performance. The experimental task will involve tracking a previously specified keyword(s) in a speech sample. The task difficulty will be manipulated by changing the number of keywords to be tracked and by changing the level of a noise masker so that the signal-to-noise-ratio changes from favorable to unfavorable. The participants will be a group of young normal-hearing individuals. Effort will be measured using several physiological markers such as galvanic skin response, heart rate, and blood oxygen level. Subjective ratings of listening effort will also be collected. For each level of difficulty, the percent correct detection performance will be calculated. Conditions that result in equal performance (for example, greater than 90% scores) will be selected for analyses. Any difference in physiological measures between these conditions will then likely reflect differences in listening effort. The study will provide preliminary candidates for developing an objective and real-time measure of listening effort.

1pSC2. Semiautomatic measurement of speech error data recorded with ultrasound. Stefan Frisch (Dept. of Commun. Sci. and Disord., Univ. of South Florida, 4202 E Fowler Ave., PCD1017, Tampa, FL 3362, frisch@cas.usf.edu)

Ultrasound has been shown to be a useful articulatory research tool. However, ultrasound imaging produces video data that is time-consuming to analyze. In this study, previous work on imaging of articulatory errors between velar and alveolar stops is extended to a larger group of participants through the use of measures extracted from semiautomatically created tongue contours that were generated using EDGETRAK software. Speech errors were induced in word onset consonants using nonsense tongue twisters. Consistent with previous research, onset stop errors were quantified using raising of the tongue dorsum toward the palate and angle of elevation of the tongue blade toward the alveolar ridge. In the majority of cases, it was found that errors created productions that were statistically indistinguishable from normal productions of the wrong category. However, some intermediate or gradient errors between categories were also found. The semiautomated measurement technique appears to be an efficient alternative to manual measurement of the data, producing qualitatively similar results. [This work was supported by NIH R03 DC06164.]

1pSC3. Gradient measures of children's speech production: Visual analog scale and equal appearing interval scale measures of fricative goodness. Kari Urberg-Carlson, Benjamin Munson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, Minneapolis, MN 55455, urbe001@umn.edu), and Eden Kaiser (Univ. of Minnesota, Minneapolis, MN 55455)

Cross-sectional and longitudinal studies of speech-sound development utilizing phonetic transcription often show an apparently discontinuous acquisition of speech sounds. Instrumental studies, in contrast, suggest that children acquire speech-sound contrasts gradually. An ideal perceptual-rating measure of children's production should capture this gradient change in production, correlate well with acoustic data, and be fast and easy both to administer and to score. A previous experiment in our laboratory compared three methods of collecting continuous judgments of children's consonant productions of /s/ and /ʃ/: Of the three, visual analog scaling (VAS) correlated best with acoustic parameters. However, as was administered in that experiment, VAS allowed only two-way comparisons between phonemes. To determine whether VAS could be used for multiple-way comparisons, two experiments were conducted. In the first, participants first made a forced choice between two phonemes, then used VAS to make a judgment of the goodness of the phoneme, from perfect to bad. A second experiment was conducted using a seven-point EAI scale in place of the VAS. The results will be correlated with acoustic data as in the previous experiment and will further our knowledge of optimal methods for measuring children's speech production perceptually.

1pSC4. Voice quality contour measurements for prosody detection. Jeung-Yoon Choi (Dept. of Elect. and Elect. Eng., Yonsei Univ., 134 Shinchon-dong, Seodaemun-gu, Seoul, Korea 120-749)

Voice quality contour measurements were used to detect prosody markers. This work is an extension of an earlier paper that examined static voice quality measures for prosody detection. Voice quality measures include those related to the voice source, such as open quotient, spectral tilt, and harmonic component distributions. Specifically, the measurements comprise: the relative amplitudes of the first two harmonics; the relative amplitudes of the first harmonic in relation to the amplitudes of the first and third formants. Contour measurements over several syllables, in addition to static measurements at the nucleus of each syllable, were used to detect ToBI markers for accent and boundaries. Experimental results show varying improvements over using only static measurements.

1pSC5. An algorithm for token-by-token linear predictive coding analysis of oral monophthongs. Leonardo Oliveira and Pablo Arantes (Dept. of Linguist., Campinas State Univ., PO Box 6045, Campinas, SP, Brazil 13083-970, leo.oliveira@gmail.com)

Linear Predictive Coding (LPC) is the most reliable vowel formant estimation method. LPC trajectories closely approximate formants and are preferable as a means of avoiding human bias. Nevertheless, formant analysis by LPC is not error free, as it can be affected by fundamental frequency, noise and vocal tract morphology. Error minimization is usually achieved through manual checking of LPC results with spectral contour. An algorithm to find optimal LPC parameters (Burg method) on a token by token basis was developed and implemented as a Praat script. The algorithm minimizes inter- and intra-vowel formant variability without knowledge of vowel identity. Algorithm results were compared with results of manual adjustment of LPC parameters on a trial and error basis on a corpus of 180 stressed and unstressed oral monophthongs of Brazilian Portuguese from a male speaker. First and second formants linear correlations are $r^2 = 0.98$ and 0.99 respectively. The areas of /i, a, u/ vocalic triangles (mean values as vertices) of manual and automatic measurements showed only 4.5% difference. These

results suggest the algorithm can be generalized to any speaker and speed up the measurement of vowel formants in large corpora. [Work supported by CNPq and Fapesp.]

1pSC6. Advanced time-frequency displays applied to forensic speaker identification. Sean A. Fulop (Linguist. PB92, Cal. State Univ. Fresno, 5245 N. Backer Ave., Fresno, CA 93740, sfulop@csufresno.edu) and Sandra Ferrari Disner (Univ. of Southern California, Los Angeles, CA 90089)

Forensic phoneticians have long utilized conventional spectrograms as the basis of the the auditory-acoustic method of speaker identification. Recent investigations [S. A. Fulop and K. Fitz, "Using the reassigned spectrogram to obtain a voiceprint," J. Acoust. Soc. Am. **119**, 3337 (2006); S. A. Fulop and S. Disner, "The reassigned spectrogram as a tool for voice identification," Int. Congress of Phonetic Sciences **XVI**, 1853–1856 (2007)] have suggested that a newer technology, i.e., the reassigned spectrogram, may be at least as valuable in the service of speaker identification. The present paper describes methods for forensic speaker identification using enhanced time-frequency displays including reassigned spectrograms and Zhao-Atlas-Marks distributions. Identification methods which employ the aforementioned displays are compared with the more established auditory-acoustic method, using recordings of 10 speakers saying several short phrases two times each. The results point to the efficacy of these newer time-frequency displays for speaker identification, particularly when used to corroborate conventional methods. One important development is the ability to focus on the detailed time-frequency features of the phonation process, thereby obtaining a new speech biometric indicative of vocal cord dynamics.

1pSC7. Automatic acquisition device identification from speech recordings. Daniel Garcia-Romero and Carol Y. Espy-Wilson (Dept. of Elec. and Comput. Eng., Univ. of Maryland, A.V.W. Bldg. 3180, College Park, MD 20742)

In the field of speaker recognition (SR), the artifacts left in the speech signal by the acquisition device are highly detrimental. When the speech signals used in the enrollment and testing phases come from different devices, the recognition accuracy decreases significantly. Hence, most SR systems try to remove these artifacts by using some general knowledge about the acquisition device. However, the degree of specificity used in the characterization of the devices is very broad (e.g., GSM handset as opposed to CDMA). This study focuses on the dual problem with a higher degree of specificity. That is, given a speech recording, how to automatically determine the exact make and model of the acquisition device regardless of the speaker and speech content. For this purpose, the HTIMIT and LLHDB databases, comprising short sentences from more than 300 speakers over 10 different acquisition devices, were used. Closed set identification experiments were conducted to assess the performance of different feature sets (i.e., MFCC and linear-FCC) along with a Gaussian mixture model classifier . The results show that statistical modeling of the device characteristics in terms of MFCC provides an excellent discriminative power among the 10 devices tested. Classification accuracies higher than 90% are reported.

1pSC8. Parametric study on speaker identification biometric system using formant analysis. Binsu C. Kovoor (Dept. of Comput. Sci., Cochin Univ. of Sci. and Technol., Cochin, India 682022, binsu@cusat.ac.in), M. H. Surpiya and K. Poulose Jacob (Cochin Univ. of Sci. and Technol., Cochin, India 682022)

Speech is the most suitable bio-feature for environments where hands free is a requirement in the security system. Speaker identification is done successfully utilizing formant frequencies in the human speech. In the prototype system developed, which can be operated without the knowledge of the users, representative features are extracted and compared with speaker specific features stored in the knowledge base. The algorithm mainly consist of four modules, namely, recording of samples and building a knowledge base of the vowel sounds, building autoregressive models, recognizing vowel and thereby identifying the speaker. The important parameters to be fixed for obtaining optimum output for a given input data of the proposed model are envelope threshold and order of parametric AR model. The envelope threshold is the parameter that differentiates numerous peaks in the speech signal while an appropriate value of the order of AR model helps to compute the frequency response of the input signal accurately. The influences of various parameters have been studied and are discussed in this paper. It is expected that the results of the study will help a biometric system developer to choose appropriate value for the parameters in the algorithm for different applications.

1pSC9. Automatic analysis of vocal manifestations of psychological states. Elizabeth Rosenfeld and Jared Bernstein (Tasso Partners, 1330 Tasso St., Palo Alto, CA 94301, elizabeth@eRosenfeld.com)

Previous research identified acoustic indicators of mood and emotion in speech [I. Murray and J. Arnott, J. Acoust. Soc. of Am. **93**, 1097–1108 (1993)], suggesting that speech signals could be measured automatically to estimate the perceived mood or affect of the speaker. Forty experimental subjects produced 1584 spoken renditions of neutral sentences, either read-aloud or repeated verbatim from a spoken model. An automatic speech processing system extracted ten paralinguistic parameters from each of these spoken responses. Samples of spoken responses were then played to psychotherapists who assigned an apparent mood to each recording. Results indicate that very reliable (0.98) estimates of speaker mood can be derived from as few as 12 judgments made when listening to spoken renditions of fixed material that contains no indication of mood in the linguistic content. Furthermore, these reliable estimates can be predicted with a correlation coefficient of 0.79 from a linear combination of only three of the ten parameters: rate of speech, response latency, and maximum speech amplitude. This prediction accuracy was achieved without using any parameters based on fundamental frequency and without any sophisticated pattern matching techniques applied to the data.

1pSC10. A step in the realization of a speech recognition system based on gestural phonology and landmarks. Vikramjit Mitra (Dept. of Elec. and Comput. Eng., Univ. of Maryland, A.V.W. Bldg., #3180, College Park, MD 20742, vmitra@umd.edu), Hosung Nam (Haskins Labs., New Haven, CT 06511), and Carol Y. Espy-Wilson (Univ. of Maryland, College Park, MD 20742)

Current state-of-the-art speech recognition systems poorly perform in casual speech as they are not robust against acoustic variations. Variability is an intrinsic problem of phone-based recognition systems as phone-units may be distinctive in the cognitive domain, but their physical characteristics can vary considerably due to coarticulation. Articulatory phonology proposes that articulatory constriction gesture is an invariant action unit and speech can be decomposed into a constellation of gestures. Speech recognition studies have recently employed this framework in the sense that speech acoustic variations can be accounted for by gestural coarticulation and reduction. The VT time-functions are time-varying physical realizations of gestural constellations at distinct vocal tract sites for a given utterance. This study aims to predict VT time-functions from acoustic signals as a component model in a complete gesture-based speech recognition system. We will compare artificial neural-network (ANN) based nonlinear regression models against hierarchical support vector regression (SVR) to obtain VT time-functions from the speech signal. A single ANN can generate all 8 VT time functions, whereas SVRs generate one output at a time, necessitating use of 8 SVRs. We also explore usage of spectral parameters such as MFCC and PLPs against knowledge-based acoustic parameters (APs).

1pSC11. Discriminating dysarthria type and predicting intelligibility from amplitude modulation spectra. Susan J. LeGendre (Speech, Lang., and Hearing Sci. , Univ. of Arizona, P.O. Box 210071, 1131 E. 2nd St., Tucson, AZ 85721-0071, sjlegend@email.arizona.edu), Julie M. Liss, and Andrew J. Lotto (Arizona State Univ., Tempe, AZ 85287-0102)

Subtypes of dysarthria are often described in terms of particular disruptions in speech rhythms. In an attempt to quantify these differences, the envelope modulation spectrum (EMS) was computed on sentences read by speakers with hypokinetic, hyperkinetic, ataxic, and mixed flaccid-spastic dysarthria, and a group of normal speakers (43 speakers total). The EMS consists of the spectra of the slow-rate (up to 10 Hz) amplitude modulations of the full signal and within seven octave bands ranging in center frequency from 125–8000 Hz. From these spectra, six variables were calculated for each band relating to the peak frequency and amplitude and the relative energy above, below, and in the region of 4 Hz. According to stepwise discriminant analysis, six of these EMS variables could correctly discriminate speaker group membership with an accuracy of 84% (better performance was obtained for distinguishing individual groups). Further, 85% of the vari-

ance in the intelligibility of the dysarthric speakers (derived from a separate set of sentences) was predictable from a subset of six of these variables. These results suggest that there are reliable perturbations of speech temporal patterns associated with dysarthria subtypes and that these perturbations play an important role in intelligibility. [Work supported by NIH-NIDCD.]

1pSC12. Beyond acoustic data: Characterizing disordered speech using direct articulatory evidence from real time imaging. Edward Holsinger (Dept. of Linguist., Univ. of Southern California, GFS 301, Los Angeles, CA 90089, eschlon@mac.com), Vikram Ramanarayanan, Dani Byrd, Louis Goldstein (Univ. of Southern California, Los Angeles, CA 90089), Maria Luisa Gorno Tempini (Univ. of California, San Francisco, San Francisco, CA 94117), and Shrikanth Narayanan (Univ. of Southern California, Los Angeles, CA 90089)

The traditional view of disordered speech characterizes speech errors as deletions, insertions, or transpositions of one or more phonetic symbols in sequence. Recent articulatory research, however, suggests that such a characterization of disordered speech is at best incomplete, and at worst incorrect. A consequence of abandoning the traditional view is that the value of phonetic transcription as a tool for analyzing speech errors is in question. We examine the discrepancy between the phonetic transcription and the actual articulatory events produced by an apraxic speaker whose articulation was imaged with real-time MRI with synchronous noise-canceled audio in a speak-after-me shadowing task. Results indicate that transcription was somewhat reliable between two trained transcribers, but several systematic discrepancies existed. In addition to the transcription being an incomplete record, we also found cases in which the transcription was misleading, unhelpful, or simply incorrect regarding the actual articulation produced by the speaker. Further examination of the acoustic signal revealed that the problems with the transcription cannot easily be overcome with detailed analysis of the acoustics. This study suggests that accurate characterization of disordered speech requires data of a type that cannot be provided solely by transcripts or acoustics. [Work supported by NIH]

1pSC13. Transition characteristics in speakers with dysarthria and in healthy controls: Part III: liquids and glides. Gary Weismer, Christina Kuo (Dept. Communicative Disord., UW-Madison, 1975 Willow Dr., Madison, WI 53706), and Yunjung Kim (LSU, Baton Rouge, LA 70803)

In two previous reports [Weismer, Kuo, & Chung, J. Acoust. Soc. Am. **121**, 3135 (2007); Kuo and Weismer, J. Acoust. Soc. Am. **124**, 2558 (2008)] the transition characteristics of speakers with dysarthria, secondary to Parkinson disease and amyotrophic lateral sclerosis (ALS), were compared to transition characteristics of healthy controls. The previous reports described the characteristics of obstruent-vowel [Weismer >etal/> (2007)] and diphthong [Kuo and Weismer (2008)] transitions extracted from an extended passage reading. Our hypothesis was that distributional analyses would show the CV, obstruent-vowel transitions to be less affected by the neurological disease as compared to diphthong transitions. The hypothesis was not supported by the data, which showed both types of transitions to be slower and less extensive than in healthy controls. The present report completes this analysis with documentation of transition characteristics of liquids and glides. Our expectation is that liquid and glide transition types will be slower and more extensive in talkers with dysarthria,as compared to healthy controls [Work supported by DC 003273.]

1pSC14. Acoustical analysis of vowel production in people with Parkinson's disease. Gabrielle Richard (Dept. de linguistique et de didactique des langues, Univ. du PQ a Montreal, Case postale 8888, succursale Ctr.-ville, Montreal, QC H3C 3P8, Canada, richard.gabrielle@gmail.com), Lucie Menard (Univ. du PQ a Montreal, succursale Ctr.-ville, Montreal, QC H3C 3P8, Canada), Henri Cohen (Univ. Paris-V, Edouard Vaillant 92774 Boulogne-Billancourt Cedex, France), and Emmanuelle Pourcher (Clinique Ste-Anne Memoire et Mouvement, PQ, QC G1R 3X5, Canada)

Speech is the sound wave produced by the vibration of the vocal chords finely tuned by the supraglottal articulators: the tongue, the lips, and the jaw modulate the vocal tract shape to produce sounds that are distinctive enough to be distinguished. But how can people who suffer from a motor control disease such as Parkinson's disease (PD) produce comprehensible speech? This work aims at studying the differences between speakers with PD and control speakers (C) regarding acoustical space, coarticulation, and contrastive emphasis. Eight speakers with PD and eight control speakers have been recorded in a speech production task. The Quebec French vowels /a/ and /i/ have been produced, in coarticulation with /b/ and /d/, in neutral and under emphasis contexts. The value of the first and second formants at the middle of the vowel, and the value of the second formant at the onset of the vowel have been extracted. Durations have also been extracted. The analysis showed that the coarticulation consonant has a greater impact on formantic values for the group with PD than the C group. Moreover, the formantic difference between the neutral and emphasis condition is greater for the group with PD than for the C group.

1pSC15. Articulatory modification induced by speaking rate change in Parkinson's Disease. Yunjung Kim (Dept. of Comm. Sci. and Disord., Louisiana State Univ., Baton Rouge, LA 70803, ykim6@lsu.edu) and Gary Weismer (Univ. of Wisconsin-Madison, Madison, WI 53706)

One of the most well-known treatment techniques to improve speech intelligibility of speakers with dysarthria involves the reduction of speaking rate (Duffy, 1995). The theory of this strategy is that a slower rate provides more time to attain articulatory targets. However, the underlying mechanism of enhanced speech intelligibility at slow speaking rate is not yet clear. In other words, it has not been directly examined whether this strategy causes measurable articulatory improvements or simply provides listeners with greater time to process impaired speech. This is an important question especially in clinical practice, related to establishing a treatment target, and furthermore, an overall treatment plan for the client. In this presentation, we will report some articulatory and acoustic indices of articulatory characteristics of speakers with Parkinson's Disease and healthy controls, producing the same utterances at three different speaking rates. [Work supported by internal funds from LSU.]

1pSC16. Acoustic characteristics of vowels in dysarthria: A comparison of clear and conversational speech. Kris Tjaden, Jennifer Lam, Joan Sussman, Stacy Blackburn, and Katrina Fulcher (Dept. of Communicative Disord. Sci., Univ. at Buffalo, 3435 Main St., Buffalo, NY 14214, tjaden @buffalo.edu)

The impact of dysarthria on vowel production has been demonstrated in a variety of studies. Acoustically, these deficits often take the form of a reduced or collapsed vowel space area, reduced F2 interquartile range, and reduced vowel transition slopes and extents. Research further suggests that these segmental deficits may help to explain reduced intelligibility or overall speech severity, at least for some speakers with dysarthria. Clear speech shows some promise for improving vowel production in dysarthria as well as for maximizing intelligibility. Comparative group studies investigating clear speech in dysarthria are lacking, however. The current study will report acoustic measures of vowel production for individuals with dysarthria secondary to Multiple Sclerosis, Parkinson's Disease, and healthy controls. As part of a larger study, speakers read 25 Harvard Sentences in a variety of speaking conditions, including Habitual or conversational and Clear. Static and dynamic segmental measures of vowel production will be reported. Group effects as well as individual speaker trends will be explored. [Work supported by NIH.]

1pSC17. The differential effects of dysarthria type on lexical segmentation. Julie M. Liss, Stephanie M. Spitzer, Kaitlin Lansford (Dept. of Speech and Hearing Sci. Arizona State Univ., Tempe, AZ 85287-0102, julie.liss@asu.edu), and John N. Caviness (Mayo Clinic Arizona, Scottsdale, AZ 85259)

Although intelligibility deficits are common across various forms of dysarthria, it was hypothesized that the causes of intelligibility reductions depend on the ways the degraded signal challenges listener cognitive-perceptual strategies. This hypothesis was confirmed by analyzing the lexical boundary errors (LBE) elicited by 44 speakers with one of four forms of dysarthria (ataxic, hypokinetic, hyperkinetic, mixed spastic-flaccid) . Error patterns from 60 listener transcripts revealed reliance on metrical stress to segment the 80 six-syllable phrases. However, reductions and disturbances in the acoustic cues to syllabic stress associated with each form of dysarthria resulted in different LBE patterns. Explanations for these unique patterns are found, in part, in the metrical structures of the dysarthric speech. For example, the temporal relationships among stressed and unstressed syllables varied widely in hyperkinetic speech, which ostensibly prevented lis-

teners from optimally exploiting metrical stress as a cue for segmentation. The equal-even temporal structure of the spastic-flaccid speech resulted in the highest proportion of boundary insertions. The presence of interword pauses facilitated the assignment of word onset to strong syllables. Results will be presented relative to models of lexical segmentation and clinical implications for targets of remediation in the dysarthrias. [Work supported from NIH-NIDCD 5R01 DC6859.]

1pSC18. Acoustic effects of familiarity on speech. Joel MacAuslan (S.T.A.R. Corp., Bedford, MA 01730, joelm@starspeech.com) and Suzanne Boyce (Univ. of Cincinnati, Cincinnati, OH 45267-0379)

In earlier work, we have reported subtle differences in acoustic measures of speech from 15 subjects who produced the Rainbow Passage at eight hour intervals over 36–42 hours of sleep deprivation. To determine whether results are due to sleep deprivation per se, or to the effect of increasing familiarity with the speech materials, we collected Rainbow Passage recordings from 15 subjects at eight hour intervals over three days while following their normal sleep/wake pattern. These data were subjected to the same measures (based on automatic detection of acoustic landmarks) as reported in our previous work. Results showed no consistent effect of familiarity with speech materials under these experimental conditions, supporting the contention that sleep deprivation affects speech articulation. [Work supported by NIH.]

1pSC19. Comparison of voice intensity effects on electroglottographic versus acoustic jitter and shimmer. Meike Brockmann (Dept. for Phoniatry/Speech Pathology, University Hospital Zurich, Frauenklinikstrasse 24, 8091 Zurich, Switzerland), Michael J. Drinnan (Newcastle General Hospital, Newcastle upon Tyne NE4 6BE, UK, M.J.Drinnan@ncl.ac.uk) and Paul N. Carding (Freeman Hospital, Newcastle upon Tyne NE7 7DN, UK, paul.carding@ncl.ac.uk)

Electroglottographic (EGG) jitter and shimmer have been described as more sensitive than acoustic perturbation. Recent studies documented significant voice intensity effects on acoustic parameters even in "normal" voice intensity. The aim of this cross-sectional single cohort study was to compare voice intensity effects in acoustic and electroglottographic measurements in several voice tasks. Forty healthy adults (1:1 f:m) aged 20–40 years phonated /a/ at subjective "soft," "medium," and "loud" voice and at 65, 75, 85, and 95 dB. Voice intensity and perturbation measurement method effects were assessed with ANOVA. Electroglottographic and acoustic perturbation were significantly different ($p<0.001$) at all intensity levels and lowest at highest voice intensities. Biggest differences were found between "soft" and "loud" voice in EGG measurements. Electroglottographic results were significantly higher in all tasks, for example mean shimmer in "soft" voice: acoustic 4.4% (SEM: 0.58%) and electroglottographic 7.35% (SEM: 1.02%). The measurement difference was smaller in prescribed intensities (mean shimmer at 65dB: acoustic 2.5% (SEM: 0.23) and electroglottographic 3.0% (SEM: 0.28%). Voice intensity changes strongly influence acoustic and electroglottographic perturbation and are minimized when patients control for voice intensity. Electroglottographic measurements are generally higher and stronger influenced by voice intensity changes. Future work should cover if this reflects a better sensitivity or more measurement error.

1pSC20. Refinement of speech motor control in preschool-aged speakers. Jennell C. Vick (Univ. of Texas at Dallas Callier Ctr. for Commun. Disord., 1966 Inwood Rd., Dallas, TX 75235, jennell@utdallas.edu), Christopher A. Moore (Natl. Inst. on Deafness and other Commun. Disord., Bethesda, MD 20892-7180), Thomas F. Campbell (Univ. of Texas at Dallas, Dallas, TX 75235), Lawrence D. Shriberg (UW-Madison, Madison, WI 53705), and Jordan R. Green (Univ. of Nebraska-Lincoln, Lincoln, NE 68583-0738)

This investigation explored whether meaningful subgroups of children would emerge from a heterogeneous population of typical preschool speakers as a result of convergences of measures from multiple speech production subsystems: behavioral, phonatory, articulatory, and respiratory. Because variability is a ubiquitous characteristic of preschool speech production, the goal of this classification was not only to minimize within-subgroup variability, but to identify the latent parameters across speech subsystems that characterized each resulting subgroup with the goal of better understanding the refinement of speech motor control that takes place at this stage of development. It is critical to understand the range of performance and interactions among speech production subsystems that exist in children with typical speech development so that we can better quantify the deviancies observed in children with speech disorders. Data were gathered from 60 speakers, aged 37–59 months, and analyzed using multivariate pattern-based classification analyses. The result was two distinct clusters of individuals whose performance across speech subsystems was differentiated by stability. One group was characterized by higher levels of variability across phonatory, articulatory, and respiratory subsystems compared to the second group. The groups did not differ significantly by composition in terms of age or sex.

1pSC21. Allophones, voice onset time, and aphasic speech errors. Adam Buchwald (Dept. of Speech-Lang Pathol. & Aud., NYU, 665 Broadway, Ste. 910, NY, NY 10012, buchwald@nyu.edu), Brenda Rapp (Johns Hopkins Univ., Baltimore, MD 21218), and Michele Miozzo (Downing St., Cambridge CB2 3EB UK)

This paper examines the speech of DLE, an aphasic English speaker who regularly deletes /s/ from /s/-stop onset clusters in English words (e.g., /spik/→ [pik]). DLE's production of stops in these words provides a window through which to examine the treatment of allophones by the speech production system. In English, the voiceless unaspirated stops that follow word-initial /s/ (e.g., the [p] in *speak*) are considered allophones of the voiceless stop (e.g., /p/), although the VOT more closely resembles that of phonologically voiced stops (e.g., /b/) which are often realized as voiceless unaspirated stops in word-initial position. Deletion errors occurring at a level before phonetic detail is generated should match the VOT of word-initial voiceless stop (e.g., peak), whereas those arising at a later phonetic level should match the VOT of the unaspirated stop (e.g., beak). In contrast to either of these possibilities, the partial activation of both aspirated and unaspirated allophones of /p/ should yield stops with a VOT that is intermediate between the voiced and voiceless phoneme productions. Thus, DLE's errors can inform our understanding of the phonetic processing of allophones in speech production.

1pSC22. Perception of talker information by children with typical and impaired linguistic development. Susannah Levi (Dept. Speech-Lang. Pathol. and Audiol., New York Univ., 665 Broadway, 9th Fl., New York, NY 10012, svlevi@nyu.edu) and Richard Schwartz (CUNY Grad Ctr., New York, NY 10016)

Despite the relatively large number of studies documenting the interaction of talker and linguistic information during speech perception in adults, such as talker-contingent phoneme processing or language-dependent processing of talker information, there is little work examining the development of talker processing in children. This study investigated the extent to which language familiarity affects the perception of the talker dimension of the speech signal by testing children's discrimination of bilingual voices across two different languages. Children—aged 7 to 12—heard pairs of words either in English (familiar language) or German (unfamiliar language) and judged whether the two words were spoken by the same or different talkers. The results of this study indicate that children are able to perform this task at above-chance levels, revealing that the perception of indexical information appears early in development. Furthermore, children are able to process this information even when they are unfamiliar with the language and, therefore, lack access to linguistic information. A comparison between children with typical language development and those with specific language impairment suggests that the perception of indexical information can be retained even when linguistic perception is impaired.

1pSC23. Non-native speech processing in children with phonological disorders. Dongsun Yim (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208, sunyim@northwestern.edu)

This study examined how children with and without phonological disorders (PD) process accented speech spoken by a non-native speaker. Two research questions were asked in this project. First was to compare how children with and without PD perform on accented speech and second was to investigate whether the exposure to the accented speech allows children's increase in overall performance. Children were aged from 3 to 8 years old who were native English speakers. The task was to identify the final words

of low- and high-context sentences spoken by two non-native Chinese speakers. Overall, it is expected that older children perform better than younger children and children perform better on high- than low-context sentences which shows children's ability to use contextual cues. For the first research question, it is hypothesized that typical children will outperform children with PD when processing accented speech which suggests unstable phonological representation in children with PD. For the second research question, it is hypothesized that children in both groups will benefit when exposed to the accented speech for the second time. However, whether children with PD will require more time to adapt to accented speech compared to typical children is of interest in the discussion.

1pSC24. Audiovisual speech perception in children with autism spectrum disorders and typical controls: Implications for development. Julia Irwin (Haskins Labs., 300 George St., New Haven, CT 06511)

For typically developing perceivers, visual speech information influences what listeners hear. For example, when the place of articulation of visual and auditory speech tokens are incongruent, perceivers often report hearing a visually influenced response (the McGurk effect; McGurk and MacDonald, 1976). To better understand the role of visual speech information in development, children with autism spectrum disorders (ASD), children with typical development and adults with typical development were assessed on: (1) sensitivity to mismatched auditory and visual speech (McGurk); (2) visual gain in the presence of auditory noise; and (3) detection of auditory and visual asynchrony. Implications for the development of audiovisual speech processing in typically developing children and adults and children with ASD will be discussed. [Work supported by NIH.]

1pSC25. Distinctive feature analysis of errant responses to Diagnostic Rhyme Test words. Edward L. Goshorn and Brett E. Kemker (Dept. of Speech and Hearing Sci., Psychoacoustics Lab., 118 College Dr. #5092, Hattiesburg, MS 39406-0001)

Ten subjects responded in a conventional multiple-choice format to two randomized lists of Diagnostic Rhyme Test (DRT) words in white noise at 3-dB signal-to-noise ratio. The DRT consists of 96 word-pairs whose initial phonemes differ by one feature: voice-unvoiced, nasal-oral, sustained-interrupted, sibilated-unsibilated, grave-acute, and compact-diffuse. List one consisted of DRT words counterbalanced for each feature and list two contained the contrasting feature. Six replicates were obtained for each subject. Errant responses were analyzed for DRT features as well as by a Chomsky-Halle feature system that permitted assigning numerical values that ranged from one to six to phonetic differences in DRT word-pairs. Results are consistent with previous findings for equality of percent correct across DRT word lists. However, DRT feature contrasts within word pairs were not equally difficult. For example, sibilated and unvoiced stimuli were significantly more difficult than their unsibilated or voiced counterpart, respectively. Also, there was no significant correlation between word error rate and the corresponding Chomsky-Halle feature value for differences in word-pairs. Thus, word difficulty appears to be unrelated to degree of feature difference in a word-pair, but rather is associated with relative phoneme audibility for the type of noise employed.

1pSC26. Context-related variability in adolescents' misarticulated and typically produced /s/. Laura L. Koenig (Haskins Labs., 300 George St., New Haven, CT 06511; Long Island Univ., One Univ. Plaza, Brooklyn, NY 11201, koenig@haskins.yale.edu) and Jonathan Preston (Haskins Labs., New Haven, CT 06511)

Acquisition of the sound /s/ is widely variable in normal development, but misarticulations that persist longer than approximately age 10 are considered to represent residual speech sound errors (RSSEs). Recent studies suggest that RSSEs may reflect subtle underlying difficulties in developing phonological representations, which, in turn, correlate with reading and spelling skills. Thus, RSSEs, even if restricted to a few sounds, may have larger clinical relevance. The acoustic characteristics of /s/ have been investigated in normally developing adolescents, but less work exists on the acoustics of misarticulated /s/, particularly considering across phonetic contexts. The current work is an exploratory study of /s/ acoustics in adolescents who do and do not misarticulate /s/. Speech samples include word lists as well as connected speech. Syllable-initial /s/ is analyzed across words varying in the following context. Of particular interest is the extent of /s/ variability as a function of the upcoming vowels and consonants (i.e., the degree of anticipatory coarticulation), and how this changes over the duration of the fricative. Instrumental data of this sort may ultimately provide greater insight into the phonological systems of adolescents with RSSEs as compared to their normally speaking peers. [Work supported by NIH.]

1pSC27. Predicting consonant recognition in quiet for listeners with normal hearing and hearing impairment using an auditory model. Tim Jürgens, Thomas Brand, and Birger Kollmeier (Medizinische Physik, Carl-von-Ossietzky Universität Oldenburg, Carl-von-Ossietzky-Str. 9-11, 26111 Oldenburg, Germany, tim.juergens@uni-oldenburg.de)

Contrary to listeners with normal hearing, listeners with high-frequency sensorineural hearing loss suffer especially from a poor consonant recognition. This is mainly attributed to the loss of audibility within high-frequency regions. In a former study [Jürgens and Brand, J. Acoust. Soc. Am. (submitted)] it was shown that consonant recognition rates in noise can be predicted well for normal-hearing listeners using an auditory model combined with a speech recognizer. The speech reception threshold (SRT) in noise could be predicted with an accuracy of 1–2 dB. In this study consonant recognition in quiet for normal-hearing and hearing-impaired listeners is investigated by regarding the individual absolute hearing threshold within the model. Observed and predicted psychometric functions of the consonant discrimination performance, consonant recognition rates of single consonants, and confusions are compared and discussed. The results show that the SRT in quiet can be predicted well for both normal-hearing and hearing-impaired listeners. Furthermore, the auditory model is modified to the signal processing of listeners with mild to moderate hearing impairment by, e.g., including a reduced dynamic compression, in order to investigate the influence on the predictions. [Work supported by DFG.]

1pSC28. Infant-directed speech to infants with a simulated hearing loss. Christa Lam and Christine Kitamura (MARCS Auditory Lab., Univ. of Western Sydney, Bankstown Campus, Locked Bag 1797, Penrith South DC NSW 1797, Australia, c.lam@uws.edu.au)

The acoustic/phonetic features of infant-directed speech (IDS) were measured during face-to-face interactions between mothers and their six-month-old normal hearing infants ($N=32$) using a setup in which the mother and infant are seated in separate rooms and interact over their respective video monitors. There were two sound conditions, (i) a *hearing condition* in which infants could hear their mothers at 55 dB SPL, and (ii) a *hearing loss condition* in which the sound was turned off, and the infants could not hear their mothers. All mothers were told *your baby cannot hear you*. Mothers and infants interacted for 10 min using three labeled toys (sheep, shoe, and shark) to facilitate the production of exemplars of the corner vowels /i/, /u/, and /a/. The results showed that mean fundamental frequency (F0), F0 range, and the duration of the IDS vowels were comparable across the two conditions. However, vowel hyperarticulation was only evident in the *hearing condition*, not in the *hearing loss condition*. These results suggest that mothers adapt their speech input according to signs of infant responsiveness. That is, when their infant cannot hear them, they maintain their modifications to IDS F0 and duration, but at the expense of vowel hyperarticulation.

1pSC29. The combined effects of distortions on vowel identification. Michelle R. Molis, Melissa A. Teahen, and Marjorie R. Leek (Natl. Ctr. for Rehab. Aud. Res., Portland VA Medical Ctr., 3710 SW US Veterans Hospital Rd., Portland, OR 97239, michelle.molis@va.gov)

The effects of intrinsic and extrinsic distortions on vowel perception were evaluated. Potential sources of distortion were high presentation level, background noise, stimulus ambiguity, and hearing status. Vowel identifications were collected for three sets of synthesized vowels that varied in both F1 and F2 (front and back vowels), or F2 and F3 (central vowels). The systematic variation in formant frequencies produced stimulus sets that contained both clear, unambiguous category tokens as well as relatively ambiguous tokens that could be identified as more than one vowel category. As overall presentation level increased from 75–90 dB SPL, identification by normally hearing listeners became less consistent for ambiguous stimuli. Additional reductions in categorization precision occurred in the presence of background noise. The effects were highly dependent on stimulus set, with the largest effects observed for the central vowels. Data from hearing-impaired listeners at high presentation levels and in noise will also be

reported. Evaluation of response patterns in the presence of combinations of these sources of distortion will help explore the difficulty people with hearing loss have processing speech in adverse listening conditions. [Work supported by OHSU Medical Research Foundation, Portland, OR.]

1pSC30. Understanding vocoded speech in steady and amplitude modulated bandpass noise. Yingjiu Nie and Peggy Nelson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, niex0008@umn.edu)

Vocoded speech perception was investigated in amplitude-modulated bandpass noise to assess normal-hearing listeners' speech understanding through cochlear implant simulations with temporal and spectral interference. Sixteen vocoder bands were applied to IEEE sentences and white noise [Fu and Nogaki (2004)]. The lower 10 bands were maintained for speech, eliminating the upper bands. The spectra of vocoded noises were set to be completely overlapping, partially overlapping, and completely separate from the speech bands. Amplitude modulations (AMs) were imposed on the noises, varying from 25–300 Hz. AM detection thresholds were measured for 80-ms white noise carriers, with rates of 25, 100, and 300 Hz. Preliminary data showed that: (1) vocoded speech perception scores were significantly different between listeners, ranging from 53–80% in quiet and from 10–54% in overlapping noise; (2) the spectra of the noise bands affected vocoder speech perception, with greater overlap causing reduced scores; (3) the noise AM rate did not significantly affect speech scores; however, vocoded speech perception in noise was significantly correlated with AM detection thresholds at 25 Hz. Results suggest that the spectral overlap of noise and speech causes reduced performance even when the noise is modulated. [Research supported by the Univ. of Minnesota.]

1pSC31. Consonant identification for hearing impaired listeners. Feipeng Li and Jont Allen (1006 N Broadway Ave., Urbana, IL 61801, fli2@illinois.edu)

This research investigates the impact of sensorineural hearing loss on consonant identification. Two elderly subjects, AS with a moderate to severe sloping hearing loss and DC with a severe sloping high-frequency hearing loss, volunteered for the pilot study. Speech perception test reveals that AS has no problem with /ta/, and has little difficulty with /pa, sa, da, za/, but never reports /ka/ and /ga/ due to a cochlear dead region from 2–3.5 kHz in the left ear, which blocks the perceptual cues for /ka/ and /ga/. In contrast, her right ear hears these sounds. Although NAL-R improves the average perception score from 0.42 to 0.53 under quiet conditions, it provides little help for noisy speech (a 3% increase at 12 dB signal to noise ratio); /ka/ and /ga/ remain unintelligible with NAL-R. The other subject DC can hear all the 16 consonants tested in both ears with the assistance of NAL-R. However, it only improves the recognition scores of low and midfrequency sounds such as /pa, ba/ and /ka, ga/. It degrades the high-frequency sounds such as /ta/ and /da/. No cochlear dead region is confirmed for DC.

1pSC32. Spectral differences between live and replayed speech measured in occluded ear. Dragana Barac-Cikoja and Jose N. Reyes (Gallaudet Res. Inst., Gallaudet Univ., 800 Florida Ave. NE, Washington, DC 20002)

To further our understanding of the transfer function between the mouth and the eardrum for occluded ears, we compared the recordings of live and played-back self-generated speech captured in the speaker's ear canals. The airborne speech feedback (during the speech production) and its recording (during the playback phase) were delivered binaurally via earmolds or insert earphones (ER-2, Etymotic Research, Inc.). Three earmold styles were tested: Full shell with the canal portion extending either to the first or the second bend of the participant's ear canal, and a custom open-fit (Microsonic 19A). In four separate conditions, in-the-canal speech recordings were made using a pair of probe tube microphones (ER-7C, Etymotic Research, Inc.) that were inserted at a fixed distance from the eardrum. Two female speakers provided samples of four word-imbedded vowels (heed, head, hawed, who'd). The participants spoke at a comfortable speaking level, and their speech was presented at a sound pressure level calibrated to match the level at the ears. The equivalent sound level and long-term average spectra of the corresponding pairs of speech samples were then analyzed as a function of the ear occlusion type. [Support provided by the Rehabilitation Engineering Research Center (RERC) on Hearing Enhancement.]

1pSC33. Acoustic-phonetic variation in word-final alveolar consonants in speech to infants with and without hearing loss. Laura Dilley (Dept. of Commun. Disord. and Dept. of Psych., Bowling Green State Univ., 247 Health Ctr., Bowling Green, OH 43403, dilley@bgsu.edu) and Laura DeMaison (Bowling Green State Univ., Bowling Green, OH 43403)

Relatively little work has examined acoustic-phonetic variability in consonants in infant-directed speech, particularly that directed to children with hearing loss. Assimilation is a form of variability which has been particularly well studied in adult-directed speech in which a word-final alveolar consonant takes the place of articulation of a following word-initial consonant (as when /n/ in *cotton* sounds like /m/ in the phrase *cotton*[m] *balls*). Two studies examined types of acoustic-phonetic variation in assimilable environments, defined as contexts in which a word-final alveolar stop (oral or nasal) is adjacent to a word-initial labial or velar consonant, in speech directed to adults and to infants with and without hearing loss. First, waveform, spectral, and perceptual information were used to classify tokens of word-final alveolar stops into four categories: assimilated, canonical, glottalized, or deleted. Next, the second formant, F2, was measured in a subset of these tokens, to gauge the extent of place assimilation in assimilable contexts. Preliminary results show substantial phonetic and acoustic variability in assimilable environments in infant-directed speech, but the extent of this variability appears similar to that in adult-directed speech. These findings have implications for understanding language development in normal-hearing children and children with hearing loss.

1pSC34. Attention to speech sounds in normal-hearing and deaf children with cochlear implants. Derek M. Houston (Dept. of Otolaryngolgy–Head & Neck Surgery, Indiana Univ. School of Medicine, 699 West Dr., Indianapolis, IN 46202, dmhousto@indiana.edu)

Cochlear implantation (CI) is a treatment for severe-to-profound hearing loss in congenitally deaf children. Although there is very little variability in auditory thresholds among CI recipients, there is a large amount of variability in language outcomes. One possible source of variability in language outcomes is individual differences in attention and cognitive processes, which may be affected by early auditory deprivation. The effect of auditory deprivation on attention to auditory input was investigated by comparing attention to speech sounds in 23 young deaf children with CIs and 23 normal-hearing age-match controls. The deaf children received their CIs between 6 and 23 months and were tested several times during the first 1.5 years after implantation. Participants were seated on their caregiver's lap in front of a TV monitor. On each trial, they were presented with a checkerboard pattern and either silence or a repeating speech sound. Normal-hearing children looked longer at the checkerboard during the sound trials at the youngest ages whereas CI children were delayed in showing a preference for the sound trials. Moreover, performance on this task correlated with language outcome measures taken two years later, suggesting that attention to speech sounds after implantation is predictive of later language development.

1pSC35. Acoustic correlates to perceived age in American English. James D. Harnsberger (Dept. of Commun. Sci. and Disord., Univ. of Florida, Gainesville, FL 32611, jharns@ufl.edu)

An acoustic analysis of voice, articulation, and prosodic cues to perceived age was completed for a speech database consisting of 150 speakers of American English. The speakers (75 male, 75 female) were recorded representing three age groups: young (18–30), middle-aged (40–55), and old (62 years and older). The database included several types of read speech: two passages (Rainbow, Grandfather), 16 sentences incorporating a wide phonemic repertoire, three sustained vowels ([a], [i], [u]), and multiple repetitions of nonsense words. Perceptual judgments of age were collected for the sentence materials from 160 listeners in total. These perceived ages were submitted to a multiple linear regression analysis with up to 34 acoustic measures representing: voice quality (e.g., jitter, shimmer, HNR, CPP), articulation (F1+F2 vowel space size, F1 and F2 of point vowels), fundamental frequency (mean, min, max, SD), intensity (SD), speaking rate, and others. An analysis of the sustained vowel data for the females showed that 14 variables were significantly correlated with perceived age (R-squared =

0.78, $p < 0.01$), while 13 significantly correlated for the male voices (R-squared = 0.57, $p < 0.01$). Results for the other classes of speech materials will also be reported.

1pSC36. Recognition of reduced speech and use of phonetic context in listeners with age-related hearing impairment. Esther Janse (Utrecht Inst. of Linguist. OTS and Max Planck Inst. for Psycholinguistics Nijmegen, Janskerkhof 13, 3512 BL Utrecht, The Netherlands, esther.janse@mpi.nl) and Mirjam Ernestus (Radboud Univ. Nijmegen and Max Planck Inst. for Psycholinguistics, 6500 AH Nijmegen, The Netherlands)

This study investigated identification of reduced words (e.g., "onry" for "ordinary") by young listeners, elderly listeners, and young listeners in whom we simulated a hearing impairment. The main question was whether elderly listeners benefit as much as young listeners from phonetic context in recovering reduced words and whether any context difference should be attributed to hearing loss or cognitive decline. Participants were presented with excerpts from a corpus of spontaneous Dutch, consisting of either the target word in isolation or in its sentential context. The sentence context was not highly semantically/syntactically constraining. As expected, elderly listeners and young listeners with a simulated hearing impairment showed lower accuracy scores than the normal-hearing young adults. More importantly, the two hearing-impaired listener groups benefited less from context than the normal-hearing young group. Within the elderly group, hearing sensitivity also predicted one's amount of context benefit. Additionally, hearing sensitivity and age, but not a cognitive processing measure, predicted overall identification performance among the elderly listeners. In conclusion, our results suggest a double handicap for both young and old hearing-impaired listeners: they have more difficulty interpreting the acoustic signals of reduced words, and also benefit less from phonetic cues contained in surrounding words.

1pSC37. Effect of priming on age-related differences in release from informational masking. Payam Ezzatian (Dept. of Psychol., Univ. of Toronto, 3359 Mississauga Rd. North, Mississauga, ON, Canada L5L1C6, payam.ezzatian@utoronto.ca), Liang Li Kathy Pichora-Fuller, and Bruce Schneider (Univ. of Toronto, Mississauga, ON, Canada L5L1C6)

In a previous study, [Freyman *et al.* (2004)] showed that presenting listeners with all but the last word of a target nonsense sentence immediately prior to presenting the full sentence in a noisy background, produced a greater release from masking when the masker was two-talker nonsense speech than when it was speech-spectrum noise, thereby demonstrating that an auditory prime could produce a release from informational masking. In Experiment 1 of this study we showed that auditory priming produced an equivalent amount of release from informational masking in good-hearing younger and older adults. To investigate the extent to which this release from informational masking was due to the semantic content of the prime, in Experiment 2 we noise-vocoded the prime (using three bands) to remove semantic content, while retaining the prime's amplitude envelope. This manipulation eliminated any release from informational masking. In Experiment 3, when the speech masker, but not the prime was vocoded, the performance of both age groups improved equivalently. These results indicate that younger and older adults benefit equally from semantic priming, and that both age groups make equivalent use of amplitude fluctuations in a masker in an informational masking paradigm.

1pSC38. Far field performance of an anechoic chamber used for speech and hearing research. Richard Godfrey, Christopher Emrick, Allison Goff, Meghan Hiss, Anna Kiener, Katherine Lamoreau, Alexandra MacMeekin, Ashley Minton, Mallory Monjot, Heather Ridenour, Jennifer Sanders, and Dana Ulakovic (Dept. of Speech and Hearing Sci. The Ohio State Univ., 110 Carmack Rd., Columbus, OH 43210, godfrey.20@osu.edu)

The department has a 13×7×8 ft fully anechoic chamber whose commissioning records were misplaced. The graduate acoustics class measured the acoustics of the chamber from 250–2000 Hz in preparation for future research projects. The focus was on the far field and its limits (near field, and approach to wedge tips). Four groups of students made measurements along the same path. The difference in the levels at a location from that at a reference location was calculated. The mean and standard deviations of these differences were then calculated to compare these measurements with the inverse square law theory. Some measurements were also made in the near field, and closer than one-quarter wavelength from the wedge tips. The average of the means from 12–74 in. was 0.73 dB below theory. The average 95% confidence limits of these data in this range was +/− 1.1 dB. These deviations compare favorably with the acceptance criteria in the ANSI standard. The near/far field guideline also appears to be reasonable. Measurements closer than one-quarter wavelength from the wedge tips may be possible. However, further extremes characterization is needed. Envelope noise reduction should also be measured.

Session 1pSP

Signal Processing in Acoustics: Acoustic Signal Processing—Innovative Techniques for Emerging Applications

Peter Gerstoft, Chair
SIO Marine Physical Lab., Univ. of California, San Diego, La Jolla, CA 92093-0238

Chair's Introduction—1:25

Contributed Papers

1:30

1pSP1. Correlation of nonisotropically distributed ballistic scalar diffuse waves in two dimensions. Richard Weaver (Dept. of Phys., Univ. of Illinois, 1110 W. Green St., Urbana, IL), Berenice Froment, and Michel Campillo (Universit Joseph Fourier, Grenoble, France)

Theorems indicating that a fully equipartitioned random wave field will have correlations equivalent to the Greens function that would be obtained in an active measurement are now legion. Studies with seismic waves, ocean acoustics, and laboratory ultrasound have confirmed this. So motivated, seismologists have evaluated apparent seismic travel times in correlations of ambient seismic noise and constructed impressive tomographic maps of seismic wave velocity, and even detected fractional secular changes in seismic wave speed at a level of 0.0001. 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 usual wave field used in long-period seismology is due to distant ocean storms and is, even in the presence of scattering, not isotropic. Here an asymptotic evaluation of the effect, on apparent travel time in a correlation waveform, is made of nonisotropic distributions of ballistic specific intensity. It is shown that for station pairs separated by one or more wavelengths, that the effect is small. A formula is derived permitting estimations of error, and corrections to travel times. It is successfully compared to errors seen in synthetic waveforms.

1:45

1pSP2. On the precision of noise-correlation interferometry. Richard Weaver (Dept. of Phys., Univ. of Illinois, 1110 W. Green St., Urbana IL), Celine Hadziioannou, Eric Larose, and Michel Campillo (Universit Joseph Fourier, Grenoble, France)

Coda Wave Interferometry analyzes changes in diffuse waveforms when source and receiver are fixed. Changes are often due to changing acoustic wave speed, perhaps a change of temperature. Alternatively, changes can be studied in noise correlations. The advantage is in not having to wait for, or generate, a precisely identical source. But interpretation is problematic; changes in the correlation of noise fields due to an uncontrolled set of sources could be due to the sources, or to the environment. Recent studies of ambient seismic noise correlations have shown apparent stretches (potentially corresponding to wave speed slowing by parts in 10 000) over periods of weeks. Changes seen at La Reunion correlate with subsequent volcanic eruptions. Changes seen at Parkfield, California occur quickly after earthquakes and then relax back over many months. Potential for seismic monitoring motivates a study of the relative effects of source and medium changes on apparent stretching or compression. Here it is shown that long duration coda like character in a waveform permits a highly accurate estimate of stretch, an estimate that is relatively independent of changes in sources. The theory is supported by laboratory measurements and seismic data.

2:00

1pSP3. Investigation of the spatial and temporal characteristics of the New Zealand seismic noise field. Laura A. Brooks, John Townend, Yannik Behr (Inst. of Geophys., Victoria Univ. of Wellington, New Zealand), Peter Gerstoft (Scripps Inst. of Oceanogr., La Jolla, CA), and Stephen Bannister (GNS Sci., Lower Hutt, New Zealand)

The use of microseismic interferometry methods to construct tomographic models requires an understanding of the spatio-temporal characteristics of the noise field, particularly in regions such as New Zealand, where multiple potential sources exist. Directional variations in noise levels compatible with large-scale differences in sea state have previously been reported (Lin *et al.*, Geophys. J. Int., 2007). The aims of the current study are to further quantify spatial and seasonal variations in the seismic noise field in conjunction with an oceanic wave state model. This involves the analysis of a larger dataset over a longer time period, and requires more sophisticated noise-field analysis. A preliminary comparison of New Zealand and environments in which the azimuthal distribution of seismic noise sources has been investigated using beamforming is presented. The applicability of beamforming to New Zealand conditions is then discussed in the context of regional ambient noise models. Methods by which estimates of the seismic noise field's spatio-temporal characteristics can be related to contemporaneous ocean wave climate observations are also highlighted.

2:15

1pSP4. Experimental active source ocean acoustic interferometry. Laura A. Brooks (Inst. of Geophys., Victoria Univ. of Wellington, New Zealand) and Peter Gerstoft (Scripps Inst. of Oceanogr., La Jolla, CA)

Green's function approximation via ocean noise cross-correlation, referred to here as ocean acoustic interferometry, has previously been demonstrated experimentally for passive noise sources. Active sources offer the advantages of higher frequencies, controllability, and continuous monitoring. Experimental ocean acoustic interferometry is described here for two active source configurations: a source lowered vertically, and one towed horizontally. Results are compared and contrasted with cross correlations of passive noise. The results, in particular differences between the empirical Green's function estimates and simulated Green's functions, are explained with reference to theory and simulations. Approximation of direct paths is shown to be consistently good for each source configuration. Secondary (surface reflection) paths are shown to be more accurate for hydrophones with a greater horizontal separation. [Work supported by the Office of Naval Research under Grant No. N00014-05-1-0264.]

2:30

1pSP5. Geoacoustic inversion using particle swarm optimization. Geoffrey F. Edelmann and Laurie T. Fialkowski (U. S. Naval Res. Lab., 4555 Overlook Ave. SW, Code 7145, Washington, DC 20375, geoffrey.edelmann@nrl.navy.mil)

A novel optimization approach is applied to geoacoustic inversion with a towed horizontal array. Particle swarm optimization, a stochastic, population-based evolutionary algorithm, is applied to geoacoustic inversion

problems in order to estimate unknown sediment and ocean parameters. Geoacoustic inversions with a towed array are applied to both synthetic data and experimental data from MAPEX 2000. Particle swarm optimization yields quick, consistent parameter estimates. These results will be compared with previously published results using an advanced rotated coordinate simulated annealing technique. [Work supported by ONR.]

2:45

1pSP6. Experimental determination of the total scattering cross section of individual scatterers in a multiple scattering medium. Shane Walker, Philippe Roux, and William A. Kuperman (Scripps Inst. of Oceanography, 9500 Gilman Dr., La Jolla, Ca 92093-0238, shane@mpl.ucsd.edu)

A problem of interest in the study of propagation through multiple scattering media is the effective scattering cross section of the individual scattering components. In the multiple scattering regime, one challenge in particular has been to determine the scattering cross section of a component scatterer in the presence of other scatterers. We apply three different methods of experimentally measuring the total scattering cross section of a scattering object. In the first method, the scattering cross of a single scatterer is self-adaptively derived from the attenuation of the coherent field as a function of distance traveled through a multiple scattering medium comprising many identical scatterers. In the second method, the ratio of the coherent to the incoherent fields is used to infer the cross section of a single scatterer in a reverberant cavity. The third method involves a direct measurement of the scattering cross section in free space. We show that the three methods produce similar results.

3:00

1pSP7. Target detection under influence of sound wave scatterers using time reversal. Yoshiaki Tsurugaya (NEC Corp., 1-10, Nishin-cho, Fuchu, Tokyo, 183-8501), Toshiaki Kikuchi (Natl. Defense Acad., Yokosuka, Kanagawa, 238-0024 Japan), and Koichi Mizutani (Univ. of Tsukuba, Tsukuba, Ibaraki, 305-8573, Japan)

In general, it is difficult to detect the forward scattering wave from the target located between the sound source and the array. In a word, because the sound wave from the sound source is strong, the forward scattering wave from the target cannot be isolated. Moreover, when the sound wave scatterers except the target exist, the scattered sound wave acts mutually, and the target cannot be isolated. Therefore, the propagation environment is divided into the reference sound field without the target and the target sound field including the target. Next, the forward scattering wave from the target is extracted by using the signal obtained in each sound field. The time-reversal processing was performed to the extracted scattered wave, and the target position was obtained. As a result, the forward scattering wave was extracted by selecting the referring sound field and the target sound field, and the position of the target could be obtained in the environment that scatterers except target influenced. To establish this method, it is necessary to confirm that the propagation environment retains the phase conjugate.

3:15

1pSP8. Innovations detector for buried ojects. Edmund J. Sullivan (EJS Consultants, 46 Lawton Brook Ln., Portsmouth, RI 02871, paddypriest@aol.com), Ning Xiang (Rensselaer Polytechnic Inst., Troy, NY 12180-3590), and James V. Candy (Lawrence Livermore Natl. Lab., Livermore, CA 94550)

The acoustic detection of buried objects is hampered by the fact that, at frequencies high enough to resolve the object, the ground is a poor conductor of the acoustic probe signal. In order to circumvent this problem, techniques have been developed whereby high-level low frequency probe signals are used to excite resonances in the object, which, in turn, cause microscopic displacement of the ground surface that can be sensed with a laser Doppler vibrometer (LDV). Although this approach shows great promise, it is limited by speckle noise. This is a form of noise arising from the coherent property of the laser signal. In this paper we demonstrate real data results using an innovations detector, which operates by sensing the change in the statistics of the returned laser backscatter signal when a target is present. When there is no target present, the signal can be modeled by a Kalman filter, such that its innovation sequence is zero mean and white. Any deviation of the signal from this model then disrupts the innovation sequence, causing a weighted sum squared residual test to indicate the presence of a signal. Real data results are shown.

3:30—3:45 Break

3:45

1pSP9. Instantaneous frequency analysis of ultrasound Doppler signal using Bayesian probability. Asif Mehmood, Paul Goggans (Dept. of Elec. Eng., Univ. of Mississippi, University, MS 38677, asif.kyani@gmail.com), and James Sabatier (Univ. of Mississippi, University, MS 38677)

In ultrasonic Doppler vibrometry systems, acoustic returned from a moving target is frequency modulated by the components of the target's velocity in the direction of the system's co-located ultrasonic transducers. Because of this, instantaneous frequency (IF) analysis of the receiver transducer output can be used to reveal the velocities of moving components. In this paper, an IF estimation algorithm using Bayesian probability theory is presented. The acoustic returns are represented by suitable localized models that are constructed on the basis of a number of moving components of the target. Then the posterior probabilities of these localized models are computed and compared to determine which model best describes the data under observation. Once the preferred model is chosen, then its parameters are estimated. Our IF estimation method performs parameter estimation and model selection simultaneously, and extends to instantaneous estimation of these nonstationary signals. The calculations are implemented using Markov chain Monte Carlo in conjunction with thermodynamic integration with simulated annealing to draw samples from the joint posterior probability for the model and the parameters. Monte Carlo integration is then used to approximate the marginal posterior probabilities for all the parameters. The performance of this method is demonstrated by simulated and experimental results.

4:00

1pSP10. Signal processing agorithms for unmanned undersea vehicles navigation and collision avoidance. Hua Lee (Dept. of Elec. and Comp. Eng., Univ. of California, Santa Barbara, CA 93106)

In this paper, signal processing algorithms for dynamic UUV homing and docking, navigation, and collision avoidance are presented. Many conventional techniques are image-based, which involved computation-intensive image reconstruction procedures, which is ineffective in dynamic real-time environment. Therefore, it is of great importance to develop computationally effective signal processing algorithms for high-precision UUV geolocation. To optimize the applicability, the algorithms presented in this paper cover all key modalities, including both active and passive, with temporal pulse and FMCW signal formats. In terms of operating configurations, the first case is classified as the passive-receiver systems. The transmitting signal format for this case is in the form of a twin-pulse set, with opposite polarity. The bearing angle is estimated from the temporal offset of the pulses. The dual version of this configuration is in the form of an active system and the double-integration algorithm can be effectively applied to both versions. The active configuration can then be extended to incorporate the FMCW signaling format. For the FMCW case, estimation of the bearing angle and range distance can be accomplished with the Fourier transform methods. In this paper, the presentation includes theoretical analysis, signal processing algorithms, and experiments.

4:15

1pSP11. Maximizing mutual information in horizontal linear arrays. Xiaoli Zhu (Dept. of Elec. and Comput. Eng., Univ. of Massachusetts Dartmouth, 285 Old Westport Rd., North Dartmouth, MA 02747, xzhu@umassd.edu) and John R. Buck (Univ. of Massachusetts Dartmouth, North Dartmouth, MA 02747)

Estimating the bearing of a sound source using a towed horizontal array is a common array processing problem. In a passive sonar system, it is common to null sources from the forward endfire direction to eliminate the noise of the towing vessel. This research designs linear nonuniform symmetric ar-

rays of a fixed aperture for estimating the bearing of sound source. Specifically, the hydrophone spacings for a symmetric linear array are chosen to maximize the upper bound on mutual information between the true bearing and the bearing estimate in white noise. The arrays' maximizing the mutual information while nulling forward endfire look significantly different from the uniform arrays commonly used in towed systems, and also differ from arrays designed to minimize spacing redundancy. Arrays maximizing mutual information are helpful when bearing estimation is considered as a quantization problem to assign the source to the correct partition. In this approach, increasing the mutual information should reduce the probability of error (P_e) in choosing the partition containing an unknown source. Simulation results demonstrate that the optimum arrays proposed here have much lower P_es than the uniform array. [Work funded by ONR code 321US.]

4:30

1pSP12. Analysis of phased array transmitters and beamformers. Gerard Coutu, Vadim Polonichko, and Paul Wanis (Teledyne RD Instruments, 14020 Stowe Dr., Poway, CA 92064)

As phased array ultrasonic transducers become more wide spread in underwater acoustic measurement applications, a better understanding of their basic performance characteristic is of benefit. The beam patterns of a phased array can be approximated by a steady state analysis. This approach indicates the directional response of an array in response to a sinusoidal excitation. However, a steady state examination does not properly describe transient behavior of the transmit waveform and also the received signal that is subsequently beamformed. We further investigate the transient performance of the beamformer using a closed-form solution to the acoustic wave equation for a multielement ultrasonic phased array transducer. The model produces transducer beam patterns and transient beamformer output for a series of reflective targets in the far-field of the transducer. The results show a close agreement with laboratory measurements and provide a tool for further investigation of phased array transducer design and performance.

4:45

1pSP13. Mathematical characterization of the acoustic brightness and contrast control problem in complex vector space. Minho Song (Graduate School of Culture Technol., Korea Adv. Inst. of Sci. and Technol., Sci. Town, Daejon 305-701, Korea) and Yang-Hann Kim (Korea Adv. Inst. of Sci. and Technol., Sci. Town, Daejon 305-701, Korea)

The acoustic brightness and contrast control methods [J.-W. Choi and Y.-H. Kim, J. Acoust. Soc. Am. **111**, 1695 (2002)] is one of the methods which can generate higher acoustic energy in desired space by controlling multiple control sources. Since the problem is represented with complex valued multidimensional vector, the relation between input control values (magnitude and phase information of control sources) and output value (acoustic energy level in desired zone) is somewhat hard to predict intuitively. In this paper, a visualized analysis method for the acoustic brightness and contrast control problems is proposed. A geometric structure is given to the problem and acoustic variables are corresponded with geometric quantities. To give a geometric structure to acoustic brightness and contrast problem, the problem is reformulated with inner product. From the geometric structure, the acoustic brightness and contrast control problems are interpreted as a simple geometry problem. From the proposed methods, various 2D simulation results with different conditions can be interpreted with simple geometry characteristic.

5:00

1pSP14. Analysis of source denoising techniques. Chris Bahr, Tarik Yardibi, Fei Liu, and Lou Cattafesta (Mech. and Aero. Eng. Dept., Univ. of Florida, 231 MAE-A, Gainesville, FL 32611-6250, cbahr@ufl.edu)

A denoising technique for acoustic signals is analyzed. Previous work derived a methodology where three microphones are assumed to view a single signal field, each under the influence of independent channel noise. Statistical relations are used to determine the coherent acoustic field at each microphone [J. Y. Chung, J. Acoust. Soc. Am. **62**, 388–395 (1977)]. The technique is validated using a simulation of a directive acoustic source, specifically a baffled piston, with multiple microphones placed in an arc array in the far field. While the solution is analytically exact for truly incoherent noise, weakly coherent noise is shown to have a significant impact on the technique output. The sensitivity of the denoised output to signal-to-noise ratio and cross-channel noise coherence is studied. In a high-channel-count array, different microphone combinations may be used for the technique. Different microphone selection algorithms are studied to find a rule set most likely to determine the true acoustic signal strength. Outlier rejection is applied to the data and shown to improve simulation results. Additional optimization schemes are applied to the data in an attempt to find alternative methods to three-microphone denoising. [Work partially supported by GE Global Research.]

5:15

1pSP15. Speech intelligence in a room via time reversal method. Gee-Pinn Too, Yu-Hao Hsieh, Chih-Hao Chou, Bo-Hsien Wu (No. 1 Univ. Rd., Dept. of Systems and Naval Mechatronic Eng., Natl. Cheng Kung Univ., Tainan, Taiwan), and Sony Lee (Industrial Technol. Res. Inst., Tainan, Taiwan)

Time reversal method has been widely used in ultrasound and underwater communication. In the present study, this method is applied in a room to enhance the received signal by increasing the signal-to-noise ratio (SNR). A passive time reversal process allows the estimation of source signal at the original source location. Both theoretical simulations and experiments are performed in the present study. The experiments are performed in an anechoic room as well as in reverberation room. The results indicate that SNR has been greatly enhanced by this approach. Also, array length, cross correlation coefficents and other parameters are also considered and discussed in the present study.

5:30

1pSP16. Compact microphone array processing for in-vehicle voice detection applications. Elizabeth A. Hoppe and Michael J. Roan (Mech. Eng., Virginia Tech, Blacksburg, VA 24060)

The main goal of voice activity detection algorithms is to determine the presence of human voice signals in a given environment. Voice activity detection is very challenging in vehicle interiors. The main challenge in detecting the presence of voice signals in vehicles is the presence of a large number of interferers and a high background noise level. Further, many types of interferers such as tire or engine noise have signals that are highly nonstationary. In this work, compact microphone arrays mounted in various locations within a vehicle are used to extract signals from locations of interest. Experimental comparisons of the performance of several voice activity detection algorithms are made for various array configurations (including single microphone) and source extraction processing algorithms. Processing algorithms considered include blind source separation algorithms such as fastICA, transfer function based inversion methods, and both fixed and adaptive beamforming techniques. The performance of compact arrays is also compared to the performance of larger distributed microphone arrays. It is shown that the use of compact microphone arrays can significantly improve voice activity detection algorithms. The gain is quantified using receiver operating characteristic type curves plotting probability of detection vs probability of false alarm.

Session 1pUW

Underwater Acoustics: Propagation/Scattering Measurement and Applications

Brian T. Hefner, Chair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105-6698

Chair's Introduction—1:00

Contributed Papers

1:05

1pUW1. Environmental variability and acoustic uncertainty in the New England Shelfbreak Test Experiments. Heather R. Hornick and James F. Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543, hhornick @whoi.edu)

Two New England Shelfbreak Test experiments were conducted off the coast of New Jersey in the Middle Atlantic Bight, a complex and highly variable environment. Water column observations were collected from moorings, CTDs, and a Scanfish. Such extensive observations are critical to determining an accurate depiction of the sound speed field variability with which to determine the uncertainty of acoustic transmission loss models. Specific environmental parameters are identified as critical keys affecting acoustic uncertainty. By determining which keys have the most significant effect on changes to sound propagation and measuring transmission loss variability based on these parameters, the uncertainty within the region is estimated comparing several previously developed methods. Given the limitations of collecting data of adequate spatial and temporal coverage, it is necessary to determine how increased numbers of observations can reduce the acoustic uncertainty and quantify the improvement in order to balance the cost of observations with an acceptable level of uncertainty.

1:20

1pUW2. Source characterization study 2007: The three-dimensional primary field 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 (Univ. of Louisiana at Lafayette, Lafayette, LA), Joal J. Newcomb (Stennis Space Ctr., MS), James M. Stephens, and Grayson H. Rayborn (Univ. of Southern Mississippi, Hattiesburg, MS)

During September 2007 the Littoral Acoustic Demonstration Center collected acoustic and related data from three moored arrays and ship-deployed hydrophones spanning the full water column to measure the 3-D acoustic field of a seismic airgun array. A seismic source vessel shot a series of lines to give detailed angle and range information concerning the field of the primary arrival. The data were collected in the western Gulf of Mexico between the East Break and Alamos Canyon regions. Peak pressures, sound exposure levels, total energy spectra, one-third octave band, and source directivity analyses are measures used to characterize the field. Three-dimensional maps of these quantities are generated to show dependence on emission and azimuthal angles and range. Three-dimensional visualizations are being developed using a visualization cave and software for 2-D cave emulation. [Research supported by the Joint Industry Programme through the International Association of Oil and Gas Producers.]

1:35

1pUW3. Observation of channel pulse responses at long ranges in the Florida Straits. Harry A. DeFerrari, Wendy Saintval, and Jennifer Wylie (Div. of Appl. Marine Phys., Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149, hdeferrari@rsmas.miami.edu)

Acoustic propagation in shallow water is observed for transmission between a towed source (250 Hz *m*-sequence) and a bottom mounted horizontal array. A total signal processing gain of 45 dB results; 24 dB from *m*-sequence pulse compression and 21 dB from array gain. The tow is out to 60 km along a path parallel to shore. Individual and separable pulse arrivals associated with single normal modes of propagation are resolved in a parameter space of intensity, arrival time, out-of-plane angle of arrival, and Doppler. Two distinct groups of arrivals are observed, direct paths with near zero horizontal arrival angle and curved inshore, out-of-plane, paths with arrival angle of about 17 deg. A typical inshore path makes several bottom bounces both up and then down the sloping bathymetry. The inshore paths have more favorable transmission (less loss) and, at long ranges, exhibit a peculiar decrease in travel time with an increase in range. At short ranges "down" Doppler is observed, consistent with the opening range of the towed source, but at the longer ranges, "up" Doppler or closing range is observed even though the source is moving away from the receiver. A 3-D PE propagation model is applied to explain observations.

1:50

1pUW4. Observations of micromultipath formation at very long ranges in shallow water. Jennifer Wylie, Harry A. DeFerrari, and Wendy Saintval (Div. of Appl. Marine Phys., Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149, jwylie@rsmas.miami.edu)

Long- and very-long-range shallow-water transmission is observed for low-frequency *m*-sequences transmissions with center frequencies of 50, 100, and 250 Hz. The source is suspended from a ship that is holding position at each of three ranges: 10, 20, and 80 km. The receiver is a bottomed horizontal array. The array is used to resolve out-of-plane arrival angles to a small fraction of a degree by observing phase rolls of the spatial coherence function of the various arriving wavefronts. At the 10 and 20 km ranges, distinct pulse arrivals are observed in one-to-one correspondence to normal mode arrivals predicted with PE propagation models. But at the 80 km range, a large number of distinct arrivals appear with intensities or arrival times inconsistent with deterministic normal mode or ray arrivals. One possibility is that these are micromultipaths from out-of-plane paths after reflections from facets or sloping bottom contours. In fact, each of the arrivals has a unique horizontal arrival angle, and the steeper the out-of-plane angle, the later the arrival—an observation that is verified with broadband PE propagation model computations for both random faceted and gently sloping bottoms.

2:05

1pUW5. Three-dimensional modeling of acoustic propagation in the Florida Straits. Wendy Saintval, Jennifer Wylie, and Harry A. DeFerrari (Div. of Appl. Marine Phys., Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149, wsaintval@rsmas.miami.edu)

A 3-D coupled broadband parabolic equation model is used to study shallow-water propagation for a direction of transmission parallel to a sloping inshore bottom bathymetry. Three sets of arrivals result: (1) direct paths—the typical 2-D model results; (2) curved inshore paths that make several bounces first up and then down the sloping bottom; and (3) micropaths that result from reflection off random facets of the bottom around the direct paths. Archival sound speed data from the Florida Straits at the site of a propagation experiment are used as input to investigate the observed anomalously low transmission loss for some of the long-range inshore paths.

It is shown that both temporal and spatial focusing effects are at play to reduce transmission loss by as much as 15 dB for the inshore paths as compared with the corresponding direct paths having nearly the same mode characteristics. Also, the models help explain a peculiar observation for some of the longer inshore paths where the travel time actually decreases with increasing range of transmission.

2:20

1pUW6. Aspect determination using a transducer with a spiral wavefront: Basic concept and numerical modeling. Brian T. Hefner (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698) and Benjamin R. Dzikowicz (Naval Surface Warfare Ctr., Panama City Div. Code T11, Panama City, FL 32407)

A circular cylindrical acoustic transducer whose direction of motion is radial produces an outgoing signal whose wavefronts are concentric circles. A similar transducer can be designed to generate spiral wavefronts where the phase of the outgoing wave changes by 360 deg in one revolution around the transducer. If these two transducers are used in tandem as a beacon, short tone bursts fired consecutively from each transducer can be used to determine the aspect angle to a receiver relative to the beacon. This is an acoustic version of the VHF omnidirectional ranging systems currently in use in airplane navigation. There are several ways of designing a transducer to produce the spiral wavefront required for the beacon. Several of these designs have been modeled numerically to understand the details of the outgoing field and their impact on aspect determination. [Work supported by ONR.]

2:35

1pUW7. Aspect determination using a transducer with a spiral wavefront: Prototype and experimental results. Benjamin R. Dzikowicz (Naval Surface Warfare Ctr., Panama City Div. Code T11, 110 Vernon Ave. Panama City, FL 32407, benjamin.dzikowicz@navy.mil) and Brian T. Hefner (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 9815-6698)

An underwater acoustical navigational beacon is built whose operation is similar to the VHF omnidirectional ranging systems used in aircraft navigation. The beacon consists of two disk transducers whose direction of motion is radial: a circular transducer with constant radius and a spiral transducer whose radius varies linearly by one wavelength over 360 deg. To determine aspect, short tone bursts are then sent from each of the transducers. A remote hydrophone can then determine its aspect to the beacon by comparing the phase between the two transducers, with the circular transducer acting as reference. Results from tank experiments demonstrate the viability of this technique. The vertical beam patterns are also determined and compared to numerical results. In addition, other techniques, such as producing a spiral wavefront using a phased array, are also discussed. [Work sponsored by the Office of Naval Research.]

2:50

1pUW8. Three-dimensional source tracking in an uncertain environment via focalization. Dag Tollefsen (Norwegian Defence Res. Establishment (FFI), 3191 Horten, Norway) and Stan E. Dosso (Univ. of Victoria, Victoria, BC V8W 3P6, Canada)

This paper develops an approach to three-dimensional source tracking in an uncertain ocean environment using a horizontal line array (HLA). The tracking algorithm combines matched-field focalization for environmental (seabed and water column) and source-bearing model parameters with the VITERBI algorithm for range-depth estimation, and includes physical constraints on source velocity. The ability to track a source despite environmental uncertainty is examined using synthetic test cases for various track geometries and with varying degrees of prior information for environmental parameters. Performance is evaluated for a range of signal-to-noise ratios in terms of the probability of estimating a track within acceptable position /depth errors. The algorithm substantially outperforms tracking with poor environmental estimates, and generally obtains results close to those obtained with exact environmental knowledge. The approach is also applied to measured narrowband data recorded on a bottom-moored HLA in shallow water (the Barents Sea) and shown to successfully track both a towed submerged source and a surface ship in cases where simpler tracking algorithms failed.

3:05—3:20 Break

3:20

1pUW9. Underwater acoustic measurements of Doppler-shifted Bragg scattering in the vertical plane from interaction with the ocean surface. Stephen D. Lynch and Gerald L. D'Spain (Marine Physical Lab., Scripps Inst. of Oceanogr., 9500 Gilman Dr. 0701, La Jolla, CA 92093-0701)

High resolution beamformer output in the vertical plane for received fields from narrowband acoustic tones transmitted at 280, 370, 535, and 695 Hz shows evidence of Doppler-shifted Bragg scattering by the ocean surface waves. The sidebands in the beamformer output are offset in frequency by amounts corresponding to ocean surface wave frequencies, while deviations in vertical arrival angle agree with those predicted by the Bragg condition through first-order perturbation theory using measured directional surface wave spectra and measured acoustic mode wave numbers. The acoustic measurements were made by a horizontal hydrophone array on the ocean bottom in a shallow water waveguide under a variety of ocean surface conditions. [Work supported by ONR, Code 321(US).]

3:35

1pUW10. Impulse response based geo-acoustic inversion using a mobile acoustic source. Kevin D. Heaney (OASIS Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039, oceansound04@yahoo.com) and Chris Emerson (OASIS, Lexington, MA 02421)

Acoustic transmission data was collected during the Office of Naval Research Quantifying, Predicting and Exploiting Uncertainty (QPE) Pilot cruise off Taiwan (Sept. 2008). Two LFM signals were transmitted from the OASIS Mobile Acoustic Source (OMAS), which is an expendable trackable navigable mobile acoustic source paired with a sonobuys receiver for *in situ* TL measurements. The LFM transmissions were centered at 600 Hz (Q=6) and 900 Hz (Q=4.5). These sources and signals have been used in the past to estimate *in situ* transmission loss from the peak of the matched filter output. Stacking the matched-filter output as a function of range leads to a measurement of the signal evolution as a function of range, permitting qualitative geoacoustic inversion from features (number of observable bottom bounces at close range and loss of higher order bottom bounces vs range) and a quantitative time-domain based geoacoustic inversion using the parabolic equation as the forward model. The resulting sediment can then be compared with the (almost independent) measurements of transmission loss to evaluate the inversions accuracy. [Work was supported by the Office of Naval Research, Code 32OA.]

3:50

1pUW11. Calibration of a passive acoustic target scattered data from the Gulf of Maine. Elizabeth T. Küsel (NOAA/PMEL, Hatfield Marine Sci. Ctr., Oregon State Univ., Newport, OR 97365, elizabeth.kusel@noaa.gov) and Purnima Ratilal (Northeastern Univ., Boston, MA 02115)

The scattered field from long vertical cylindrical targets, such as BBN targets, can be calculated by direct application of Green's theorem and making use of appropriate boundary conditions. In a recent paper submitted to J. Acoust. Soc. Am., the authors show that the scattered levels from BBN-type targets in ocean waveguides with depth-dependent sound speed profile are influenced by the refractive nature of the incident field over the target depth extent. In the current work, the scattering model for vertically extended cylindrical (VEC) targets in range-dependent ocean waveguides is validated with experimental data from the 2006 ONR-sponsored ocean acoustic waveguide remote sensing (OAWRS) experiment in the Gulf of Maine. In order to take into account the scintillation in the measured scattered intensity caused by spatial and temporal fluctuations of the ocean waveguide, a stochastic approach is adopted. Monte Carlo simulations of the scattered field statistics are calculated by implementing the VEC model in a range-dependent environment randomized by internal waves. The simulated scattered field statistics for different source-receiver-target configurations is compared to the scattered returns from BBN targets measured during the 2006 Gulf of Maine experiment. Results show good agreement between the simulated and experimental scattered field statistics.

4:05

1pUW12. High frequency scattering from inclusions in marine sediments: Data/model comparisons. Anatoliy Ivakin (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

Recent theoretical modeling has shown that the size-distribution of inclusions in marine sediments, such as gravel-sized grains and shells embedded in sandy bottoms, is critical for high-frequency scattering. Analysis of existing data on the bottom scattering strength and sediment ground truth also shows that the backscatter data are primarily responsive to the coarse contents of the sediment. Previous ground truth data analysis, usually based on the sediment sieving, assumed that particles are nearly spherical, so that the sieve size is approximately equal to the equivalent diameter (diameter of sphere with the same volume). Coarse particles, such as shell fragments, however, appear to be very nonspherical. Therefore, a relationship between the particle sieve size and equivalent diameter, defined by the particle shape, is needed to provide required size-distributions for acoustic scattering models. Recent statistical analysis of sediment samples collected during the SAX04 experiment resulted in a set of such empiric relationships, and also provided estimating the particle shape factor, known to be an important parameter in the sediment transport processes. Obtained size distributions were used for calculations of frequency-angular dependencies of the bottom backscattering strength and provided a good agreement in the SAX04 scattering data/model comparisons. [Work supported by ONR.]

4:20

1pUW13. Source amplitude spectral information in matched-field localization. Michael J. Wilmut and Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria B.C. V8W 3P6, Canada)

This paper examines a variety of approaches to make use of knowledge of the relative amplitude spectrum of an acoustic source (but no knowledge of the phase spectrum) in multifrequency matched-field processing for source localization. A common example of this procedure involves cases where the source amplitude spectrum can be considered flat over the frequency band of interest. The primary issue is how to combine the information of complex acoustic fields at multiple frequencies, given the unknown phase spectrum. Approaches examined include maximum-likelihood phase estimation, pair-wise processing, and phase rotation to zero the phase at a specific sensor or to zero the mean phase over the array. The performance of the various approaches (processors) is quantified in terms of the probability of localizing the source within an acceptable range-depth region, as computed via Monte Carlo sampling over a large number of random realizations of noise and of environmental parameters. Processor performance is compared as a function of signal-to-noise ratio, number of frequencies, number of sensors, and number of time samples (snapshots) included in the signal averaging.

4:35

1pUW14. Mode characterization in shallow water using warping tools. Julien Bonnel, Barbara Nicolas, Jerome I. Mars (GIPSA-Lab, Dept. Images-Signal, Grenoble INP, rue de la houille blanche BP46, 38402 Saint Martin d'Hres, France, julien.bonnel@gipsa-lab.inpg.fr), Gregoire Le Touze (LMA, Aix Marseille Univ., 13402 Marseille, France), and Dominique Fattaccioli (CTSN, DGA, 83 800 Toulon, France)

In underwater acoustics, shallow water is a complex and dispersive medium. For low frequencies, the propagation is described by normal mode theory. Modes are nonlinear structures overlapped in time, in frequency, and in time-frequency domains. Advanced signal processing methods are needed to study them. This paper presents two different warping methods allowing modal separation and estimation of parameters of the modes. As these transformations are invertible, they also allow modal filtering. Thus, they are a good preprocessing tool for source localization or tomography. This is shown on simulations and real data. Warping is a transformation of a propagated signal and is based on an environment model. Here the environment is modeled with isovelocity or Pekeris waveguide. However, both methods are quite robust to mismatches with the real environment and can be used with minimal preliminary knowledge of the real environment. The first warping method is applied on the time domain. It transforms each mode into a sinus function of frequency f_m (with f_m the cutoff frequency of the mode m). The second is applied on the frequency domain. It transforms a chosen mode into a dirac, which is localized on the arrival time of the mode.

4:50

1pUW15. Warping based lag-Doppler filtering applied to motion effect compensation in acoustical multipath propagation. Nicolas Josso (GIPSA-lab, Grenoble Inst. of Technol., GIT, 961, rue de la Houille Blanche, 38402 St Martin d'Hères, France, nicolas.josso@gipsa-lab.inpg.fr), Cornel Ioana, Jérôme Mars (Grenoble Inst. of Technol., 38402 St Martin d'Hères, France), Cédric Gervaise (ENSIETA, Univ. Européenne de Bretagne, 29806 Brest Cedex, France), and Yann Stephan (SHOM, Brest Cedex 2, France)

The estimation of an impulse response of a propagation channel is necessary for a large number of underwater applications: underwater communication, sonar detection and localization, marine mammal monitoring, etc. Basically it informs us about the distortions of transmitted signal in underwater channels and enables geoacoustic inversion. The propagated signal is usually subject to additional and undesirable distortions due to the motion of the transmitter-channel-receiver configuration. This paper points out the effects of the motion while estimating the IR by matching filtering between the transmitted and the received signals. A methodology to compare the IR estimation in motionless and motion contexts is presented. Based on this comparison, a warping-based lag-Doppler filtering method in the wideband ambiguity plane for motion effect compensation is proposed. It aims at reducing the distortions due to the motion phenomena and accurately estimating the IR of the channel. The proposed methodology is applied to real data sets issued from BASE'07 campaign (NURC, SHOM, and FWG, South of Sicilia, 2007). Motion compensated estimations of IRs are computed from sources transmitting broadband linear frequency modulations moving up to 12 knots in the shallow water environment of the Malta plateau.

5:05

1pUW16. Synthesis of adaptive processing of a passive fathometer. Chris H. Harrison (NURC, Viale San Bartolomeo 400, 19126 La Spezia, Italy), James Traer, and Peter Gerstoft (Scripps Inst. of Oceanogr., UC San Diego, La Jolla, CA 92093)

By using ambient noise array-processing techniques, it has been demonstrated that a passive array can be used as a fathometer [Siderius *et al.*, JASA (2006)] and that adaptive processing gives higher temporal resolution. We show that the waveform extracted with the adaptive algorithm has a phase error of pi relative to that of conventional processing (independently observed at SIO and NURC). For the case of an array with regularly spaced hydrophone elements, the cross-spectral density (CSD) matrix is Toeplitz and the symmetry of Toeplitz matrices allows a simple analytic expression for the phase of the fathometer output. This expression is used to characterize the introduction of phase errors by adaptive processing. The effect of this phase error is investigated with simulated and real data on a vertical hydrophone array.

5:20

1pUW17. Model and measurement of high-frequency bottom backscattering intensity envelope. Jinrong Wu, Licheng Lu, Shengming Guo, and Li Ma (Inst. of Acoust., Chinese Acad. of Sci., No. 21, BeiSiHuan XiLu, Beijing, 100190, ymwjr@yahoo.com.cn)

High-frequency bottom scattering, generally, can be attributed to bottom interface roughness and volume inhomogeneities. The full grazing angle (0–90 deg) intensity envelope model of high-frequency bottom backscattering generated by these two kinds of mechanism was presented. Small-slope approximation method was used to explain bottom roughness backscattering. Bottom volume scattering coefficient was used to calculate bottom volume scattering assuming the volume scattering is homogeneous. The model incorporates the shape of transmitted pulse, directionality of transducer, scattering area, bottom characteristic, and so on. Tank experiment result shows the bottom roughness is the key factor to fine sand bottom backscattering. This backscattering intensity envelope model is useful for the application of bottom backscattering signal, such as sediment classification and geoacoustic inversion. [Work supported by the NSFC.]

Meeting of Accredited Standards Committee (ASC) S2 Mechanical Vibration and Shock

A. T. Herfat, Chair S2
Emerson Climate Technologies, Inc., 1675 W. Campbell Rd., P. O. Box 669, Sidney, OH 45365-0669

R. L. Eshleman, Vice Chair S2
Vibration Institute, 6262 Kingery Hwy., Ste. 212, Willowbrook, IL 60527

Accredited Standards Committee S2 on Mechanical Vibration and Shock. Working group chairs will report on the status of various shock and vibration standards currently under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAG for ISO/TC 108, Mechanical vibration, shock and condition monitoring, and its five subcommittees, take note - that meeting will be held in conjunction with the Standards Plenary meeting at 9:00 a.m. on Tuesday, 19 May 2009.

Scope of S2: Standards, specification, methods of measurement and test, and terminology in the field of mechanical vibration and shock, and condition monitoring and diagnostics of machines, including the effects of exposure to mechanical vibration and shock on humans, including those aspects which pertain to biological safety, tolerance and comfort.

MONDAY EVENING, 18 MAY 2009 NEWMARK THEATER, 7:00 TO 9:00 P.M.

Session 1eID

Interdisciplinary: Tutorial Lecture: The Art and Science of Unique Musical Instruments

Uwe J. Hansen, Chair
Dept. of Chemistry and Physics, Indiana State Univ., Terre Haute, IN 47809

A tutorial presentation on The Art and Science of Unique Musical Instruments will be given by Ela Lamblin on Monday, 18 May, at 7:00 p.m.

Ela Lamblin has created many unique musical instruments and sound sculptures in the Northwest. Together with his partner and wife, Leah Mann, he founded the performance company Lelavision. Ela and Leah will demonstrate a variety of instruments, and with the help of ASA's acousticians, explain the sounds they produce—some of which are not what we usually hear from traditional instruments. For a preview visit www.lelavision.com/

The presentation and performance will be held in the modern Newmark Theater (www.pcpa.com/events/newmark.php), located two blocks from the Hilton Hotel in Antoinette Hatfield Hall, 1111 SW Broadway (at Main Street).

Session 2aAAa

Architectural Acoustics: Multiple Channel Systems in Room Acoustics

Ning Xiang, Cochair
Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180

Boaz Rafaely, Cochair
Electrical and Computer Engineering Dept., Ben Gurion Univ., Beer Sheva, Israel

Chair's Introduction—8:00

Invited Papers

8:05

2aAAa1. A multi-channel audio system based on the theory of integral equations. Filippo M. Fazi and Philip A. Nelson (Inst. of Sound and Vib. Res., Univ. of Southampton, Highfield, S0171BJ, Southampton, U.K., ff1@isvr.soton.ac.uk)

The basics of a multi-channel audio system, which attempts the reproduction of a desired sound field, are presented. The system's hardware consists of a three-dimensional array of loudspeakers, and can be used in combination with a specially designed microphone array. The mathematical fundamentals on which this technique is grounded consist of the formulation of the problems as an integral equation. The loudspeaker signals are determined from the knowledge of the target sound field on the boundary of a given control volume. The solution to this inverse problem is computed performing a singular value decomposition of the integral operator involved. For some simple array geometries it is possible to calculate an analytical solution to the problem. A regularization method is applied, as required by the ill-posed nature of the inverse problem under consideration. Some insight into the physical meaning of the ill-posedness is given and some analogies to near-field acoustic holography are suggested. The effectiveness of the method proposed has been verified experimentally and some of the experimental results are presented. Finally, it is shown how this technique has been successfully applied to the design of a multi-channel auralization system for room acoustics.

8:25

2aAAa2. Representation of the musical instruments directivity using dodecahedron loudspeakers. Gottfried Behler and Martin Pollow (Inst. of Tech. Acoust., RWTH Aachen Univ., 52056 Aachen, Germany, gkb@akustik.rwth-aachen.de)

Room-acoustical measurements in general are performed with omnidirectional sound sources. With respect to auralization, such an impulse response may not be ideal since it does not represent the situation playing an instrument in the room. To achieve the directivity of a real source (such as an instrument or human voice) with a technical sound source (a loudspeaker) requires either to copy the body and the surface velocity distribution of that particular source or to reproduce the directional pattern of the radiation using a multiple source configuration like a dodecahedron loudspeaker array with independent excitation of each transducer. The advantage of the latter method is obvious since one single source is able to provide different directivities by changing the excitation profile. To maintain the appropriate excitation of each individual transducer, different approaches can be made. The method described here uses spherical harmonics decomposition of the target radiation pattern and a subsequent calculation of the frequency dependent excitation coefficients for each transducer. The advantage of this method is a flexible and fast calculation delivering filters that can be used either for real time convolution or off-line processing of the measuring signals. To measure the instruments directivity an array with 32 microphones is used.

8:45

2aAAa3. Considering modal aliasing in the implementation of an acoustic echo canceller in the wave domain. Martin Schneider and Walter Kellermann (Multimedia Commun.s and Signal Processing, Univ. Erlangen-Nuremberg, Cauerrstr 7, 91058 Erlangen, Germany, schneider,wk@LNT.de)

Traditional multichannel acoustic echo cancellation applied to wave field synthesis systems is computationally extremely expensive due to the large number of channels to be identified. By using wave domain adaptive filtering (WDAF) one is able to choose an approximation which neglects the interaction (exchange of energy) between different modes. This reduces the number of necessary adaptive filter coefficients drastically. Furthermore, it allows to use computationally relatively inexpensive single-channel filter adaptation algorithms. Then, however, modal aliasing due to the limited resolution of the used microphone array becomes a significant problem since it constitutes an interaction between different modes with respect to the microphone signals. The effects of modal aliasing on the adaptation behavior of a WDAF-based echo cancellation system are analyzed and a modified wave domain representation of the synthesized wave field is proposed which reflects the limited resolution of the microphone array.

9:05

2aAAa4. Spherical harmonic beamforming for room acoustic analysis. Gary W. Elko and Jens Meyer (mh acoustics LLC, 25A Summit Ave., Summit, NJ 07901)

We describe the potential of using a spherical beamforming microphone array to investigate the spatial correlation of sound fields in rooms. We are building spherical microphone arrays consisting of many acoustic pressure sensors mounted appropriately on the surface of a rigid sphere. Our associated spherical eigenbeamformer decomposes the sound-field into spatially orthonormal spherical harmonics up to third-order. We refer to the signals from the eigenbeamformer as eigenbeams. All eigenbeams have phase centers at the physical center of the array. Due to the orthonormal property of the eigenbeamformer, a diffuse field ideally results in zero correlation between the eigenbeams. Therefore, by measuring the cross-correlation between the eigenbeam signals, one can investigate the proximity to the isotropy (or "diffuseness") of the sound-field. Simultaneously, the underlying eigen-beam patterns can be steered without effecting the orthonormality property. How the cross-correlation function changes with general orientation of the eigenbeams is another potential measure for sound field diffuseness in rooms. We will show some real room measurements demonstrating the potential usefulness of this approach.

9:25

2aAAa5. Imaging room acoustics with the audio camera. Adam O'Donovan, Ramani Duraiswami, Nail A. Gumerov, and Dmitry N. Zotkin (Perceptual Interfaces and Reality Lab., Dept. of Comput. Sci. Univ. of Maryland, College Park, MD 20742)

Using a spherical microphone array and real time signal processing using a graphical processing unit (GPU), an audio camera has been developed. This device provides images of the intensity of the sound field arriving at a point from a specified direction to the spherical array. Real-time performance is achieved via use of GPUs. The intensity can be displayed integrated over the whole frequency band of the array, or in false color, with different frequency bands mapped to different color bands. The resulting audio camera may be combined with video cameras to achieve multimodal scene capture and analysis. A theory of registration of audio camera images with video camera images is developed, and joint analysis of audio and video images performed. An interesting application of the audio camera is the imaging of concert hall acoustics. The individual reflections that constitute the impulse response measured at a particular seat may be imaged, and their spatial origin determined. Other applications of the audio camera to people tracking, noise suppression, and camera pointing are also presented. [Work partially supported by NVIDIA and the VA.]

9:45

2aAAa6. Proposed method to measure the diffusion coefficient. Peter D'Antonio (651-C Commerce Dr., Upper Marlboro, MD 20774, pdantonio@rpginc.com)

A method to measure the uniform diffusion coefficient has been published as an AES Information Document [AES-4id-2001, JAES, Vol. **9**, pp. 148–165 (March 2001)]. The method utilizes 37 fixed pressure zone microphones separated by 5 degrees located on a 1 m semicircle and a loudspeaker located on a 2 m concentric semicircle. A sample array sits at the origin. Under computer control 37 impulse responses are collected for the sample, background without sample, and a reference flat panel of similar size. The data reduction procedure will be reviewed to include all developments since the 2001 publication. In addition, experimental data will be presented along with the current presentation format. Currently the method is in the final stages of review for incorporation as Part 2 of ISO 17497-1, which describes the procedure to measure the random incidence scattering coefficient. A review and comparison between the scattering and diffusion coefficients will also be presented. A summary of this research can be found in the reference text [T. J. Cox and P. D'Antonio, Acoustic Absorbers and Diffusers: Theory, Design and Application, Spon Press 2004, second edition in press].

10:05—10:15 Break

10:15

2aAAa7. Analysis of *in situ* acoustic absorption using a spherical microphone array. Jonathan Rathsam and Boaz Rafaely (Dept. of Elec. and Comp. Eng., Ben-Gurion Univ. of the Negev, Beer-Sheva, Israel 84105, jrathsam@gmail.com)

Acoustic absorption of materials typically depends on the characteristics of the incident sound field. Differences in the incident sound field between one laboratory and another, or between the laboratory and *in situ*, may therefore lead to differences in the measured absorption for the same material. To overcome this problem, a new method is proposed to measure acoustical absorption using a spherical microphone array. With this method one measures simultaneously the acoustic absorption as well as the incident sound field. The incident sound field and the reflected sound field are separated by the spherical microphone array using spatial filtering, or beamforming. The acoustic absorption is then determined from the amplitudes of the incident and the reflected fields. The measurement method may also be useful for verifying sound field diffusivity in a reverberation chamber when measuring diffuse-field absorption. Furthermore, absorption coefficients measured under in-situ sound fields may yield more accurate computer-based sound field simulations than diffuse-field absorption coefficients when the in-situ sound field is not diffuse. [Work supported by the National Science Foundation.]

10:35

2aAAa8. Speech intelligibility improvement in rooms using microphone arrays. Michael R. Stinson (Inst. for Microstructural Sci., Natl. Res. Council of Canada, Ottawa, ON K1A 0R6 Canada), Bradford N. Gover (Inst. for Res. in Construction, Natl. Res. Council of Canada, Ottawa, ON K1A 0R6 Canada), and Gilles A. Daigle (Natl. Res. Council of Canada, Ottawa, ON K1A 0R6 Canada)

Microphones arrays reduce the effects of reverberation and background noise as compared to a single microphone. The resulting improvement in speech intelligibility in rooms is investigated here, following the approach discussed by Houtgast *et al.* [Acustica **46**, 60–72 (1980)]. Room characteristics (volume, reverberation time), the background noise spectrum, and the source-receiver separation are used to compute a modulation transfer function. This function is input into the procedure described in IEC standard 60268-16(2003)

to arrive at an objective measure of speech intelligibility, the speech transmission index (STI). Doing this for both omnidirectional and microphone array receivers, the improvement in intelligibility is determined. The signal source has the spectral shape of a human talker, as given in the IEC standard; the background noise spectrum is based on measurements made in a restaurant. The benefit of using microphone arrays is discussed for various conditions.

10:55

2aAAa9. Inter-aural cross correlation in a sound field represented by spherical harmonics. Boaz Rafaely and Amir Avni (Dept. of Elec. and Comput. Eng., Ben-Gurion Univ. of the Negev, Beer Sheva 84105, Israel)

Spaciousness is an important acoustic feature of concert halls. An accepted measure for spaciousness is the inter-aural cross-correlation (IACC), which employs the transfer function from a source in the hall to the two ears of a listener, the latter referred to as head-related transfer functions (HRTF). Recently, spherical microphone arrays have been studied for room acoustics analysis and music recordings. As these arrays typically use a finite number of microphones, they may not be able to capture the spatial information required for complete spatial analysis or for sound reproduction with realistic spaciousness. This study proposes the use of spherical harmonics representations for both the HRTF data and the sound field data, facilitating IACC analysis for sound fields represented by a finite order in the spherical harmonics domain. The effect of limited spherical harmonics order on the spaciousness of various sound fields is finally studied using simulated and measured data.

Contributed Papers

11:15

2aAAa10. A new method for the measurement of directional sound arrivals in rooms. Brian T. Thornock (Acoust. Res. Group, Dept. of Phys., Brigham Young Univ., N-283 ESC, Provo, UT 84602, brianthornock@gmail.com) and Timothy W. Leishman (Brigham Young Univ., Provo, UT 84602)

In room acoustics, directional measurements have the ability to more fully detail room effects than common nondirectional measurements. The extraction of directional information has been a subject of considerable interest for years and has resulted in the development of several interesting approaches to the problem. However, past methods have often suffered from inaccuracies, difficult implementation, lengthy procedures, extensive computational requirements, or the need for large amounts of equipment. A practical new method will be presented that extracts directional information of sound arrivals through a single multichannel measurement and a new cross-correlation technique. A direct comparison of numerical simulations will demonstrate that the method is more accurate than other well-established methods. Advantages and limitations of the new approach will be discussed.

11:30

2aAAa11. Microphone array beamforming with near-field correlated sources. Jonathan L. Odom (Dept. of Elec. and Comput. Eng., Duke Univ., CIEMAS 2410B, Durham, NC 27708, jlo11@duke.edu) and Jeffery Krolik (Duke Univ., Durham, NC 27708)

As the price of omnidirectional microphone arrays has fallen, new applications have emerged for large room audio capture. Further, large room acoustics use loudspeakers, which create correlated sources and severely limit the use of optimum beamformers. Most sources are in the near field due to the large wavelength of speech. Widrow and Kailath's work on correlation in sonar and radar assumed far field, but it cannot be directly applied to near-field acoustics. A new beamforming method has been developed, which incorporates near field and correlation. The source bearings are estimated and an uncorrelated model is formed. A better DOA estimator was developed for the near-field correlated case limited by the width of the mainlobe. The new beamformer is able to null strong, perfectly and nearly perfectly correlated signals using the uncorrelated model and minimum variance distortionless response beamforming.

11:45

2aAAa12. Prediction and measurement of reverberation increase from electroacoustic architecture systems. Roger Schwenke (Meyer Sound Labs., 2832 San Pablo Ave., Berkeley, CA 94702, rogers@meyersound.com), Steve Ellison (Meyer Sound Labs., Sierra Madre, CA 91024), and Mark Poletti (Industrial Res. Ltd., Lower Hutt, New Zealand)

Electroacoustic architecture (EA) systems increase the reverberation time (RT) of rooms by regenerating the physical reverberation of the room, and/or adding electronic reverberation to the direct sound. In regenerative systems, the amount of regeneration is controlled by the loop gain of the system. Increasing loop gain is analogous to reducing the effective absorption in the room. The RT of the electronic reverberator is analogous to the decay time of a physically coupled room, and, therefore, analogous to adding cubic volume to the room. As the electronic RT is increased, there is a point above which the resulting reverberation will begin to be double sloped. This point depends on the gain. The higher the gain, the higher the electronic RT before the resulting reverberation double slopes. In previous work, the gain before feedback of EA systems has been predicted in a statistical approach based on the bulk properties of the room such as RT and cubic volume. This, in turn, can be used to predict the maximum possible power gain and RT increase before double sloping. In this paper, we measure systems in use in the field and compare the achieved performance with predictions.

Session 2aAAb

Architectural Acoustics and the National Council of Acoustical Consultants: Student Design Competition

Robert C. Coffeen, Chair
School of Architecture and Urban Design, Univ. of Kansas, Lawrence KS 66045

The Technical Committee on Architectural Acoustics of the Acoustical Society of America, the National Council of Acoustical Consultants, and the Robert Bradford Newman Student Award Fund are sponsoring the 2009 Student Design Competition that will be professionally judged at this meeting. The purpose of this design competition is to encourage students enrolled in architecture, engineering, physics, and other university curriculums that involve building design and/or acoustics to express their knowledge of architectural acoustics and noise control in the design of a facility in which acoustical considerations are of significant importance.

This competition is open to undergraduate and graduate students from all nations. Submissions will be poster presentations that demonstrate room acoustics, noise control, and acoustic isolation techniques in building planning and room design. The submitted designs will be displayed in this session and they will be judged by the panel of professional architects and acoustical consultants. An award of USD $1250.00 will be made to the entry judged "First Honors." Four awards of USD $700.00 will be made to each of the four entries judged "Commendation."

Session 2aAB

Animal Bioacoustics and Acoustical Oceanography: Autonomous Remote Monitoring Systems for Marine Animals I

Marc O. Lammers, Chair
Hawaii Inst. of Marine Biology, Kailua, HI 96734

Chair's Introduction—8:00

Invited Papers

8:05

2aAB1. A survey of remote underwater autonomous acoustic recorders. Whitlow W. L. Au, Marc O. Lammers (Hawaii Inst. of Marine Biology, Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734), and Michael F. Richlen (Univ. of Hawaii Honolulu, HI 96734)

There has been a proliferation of different types of remote underwater autonomous acoustic recorders in recent years. Many were designed for a specific purpose to measure sounds from specific organisms. Others were designed for general application to encompass a wide variety of sounds. Many have fancy acronyms such as ARP, BPROBE, DTAG, EAR, EARS, HARP, MARU, PALs, RUDAR, and TPOD, to name a few. All have essentially the same subsystems consisting of a hydrophone and amplifier assembly, an analog-to-digital converter, data storage, and a microcontroller to handle the data acquisition and storage operations and to control the overall functioning of the unit. The differences are in the bandwidth of the system, amount of storage available, amount of power required, and duration of deployment possible. Some can be purchased commercially, some can only be leased, while others can only be used in a collaborative research project. The prices also vary considerably. Some have been used for many years while others have come into play within the last few years. The sum total of all these acoustic recorders have contributed tremendously to our understanding of the marine ecosystem in ways that we could not have imagined several years ago.

8:25

2aAB2. Results and insights from operational acoustic monitoring networks. Christopher W. Clark (Bioacoust. Res. Prog., Cornell Univ., 159 Sapsucker Woods Rd., Ithaca, NY 148504), William T. Ellison (Marine Acoust., Inc., Middletown, RI 02842), Brandon L. Southall (Natl. Marine Fisheries Svce., Silver Spring, MD 20910-6233), Adam Frankel (Marine Acoust., Inc., Middletown, RI 02842), Dimitri Ponirakis, Christopher Tremblay, Ann Warde, and Eric Spaulding (Cornell Univ., Ithaca, NY 148504)

Data from two acoustic monitoring networks operating off New England in an area frequented by whales reveal acoustic features of those habitats. These seafloor and moored systems continuously sample the acoustic environment, and resultant data provide mechanisms for mapping, quantifying, and describing the spatio-spectral-temporal variability of the acoustic habitat over ecologically meaningful scales. By focusing on species-specific frequency bands used by fin, humpback, and right whales for long-range communication, we are beginning to measure the acoustic dynamics of their primary communication channels. Results reveal the extent to which different sources of sound in the ocean, both natural and man-made, influence the probability of whale communication. In some habitats with high rates of vessel traffic and high levels of vessel noise, the predicted area over which animals can communicate is reduced to a small proportion of what it would be under quiet conditions. The dynamics of this masking effect are highly variable and vary by species. When considered from a large-scale and behavioral ecological perspective, this reduction in acoustic habitat, as measured in terms of the proportional loss of communication space, likely represents a significant loss for species in which acoustic communication is known to serve critical biological functions.

Contributed Paper

8:45

2aAB3. PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans. Douglas Gillespie (Sea Mammal Res. Unit, Univ. of St. Andrews, KY16 8lB, Scotland, UK), David K. Mellinger (Oregon State Univ. and NOAA Pacific Marine Environ. Lab., Newport, OR 97365), Jonathan Gordon (Ecologic UK, Newport on Tay, Fife, DD6 8JH, UK), David McLaren, Paul Redmond, Ronald McHugh, Philip Trinder, Xiao-Yan Deng (Heriot-Watt Univ., Edinburgh, Scotland, UK), and Aaron Thode (Univ. of California, San Diego, La Jolla, CA 92093)

PAMGUARD is open-source, platform-independent software to address the needs of developers and users of Passive Acoustic Monitoring (PAM) systems. For the PAM operator—marine mammal biologist, manager, or mitigator—PAMGUARD provides a flexible and easy-to-use suite of detection, localization, data management, and display modules. These provide a standard interface across different platforms with the flexibility to allow multiple detectors to be added, removed, and configured according to the species of interest and the hardware configuration on a particular project. For developers of PAM systems, an Application Programming Interface (API) has been developed which contains standard classes for the efficient handling of many types of data, interfaces to acquisition hardware and to databases, and a GUI framework for data display. PAMGUARD replicates and exceeds the capabilities of earlier real time monitoring programs such as the IFAW Logger Suite and Ishmael. Ongoing developments include improved real-time location and automated species classification. [PAMGUARD funded by the OGP E&P Sound and Marine Life project.]

Invited Papers

9:00

2aAB4. Software for bioacoustic analysis of passive acoustic data. David K. Mellinger (Cooperative Inst. for Marine Res. Studies, Oregon State Univ. and NOAA Pacific Marine Environ. Lab., 2030 SE Marine Sci. Dr., Newport, OR 97365), Douglas Gillespie (Univ. of St. Andrews, KY16 8lB, Scotland, UK), Harold Figueroa (Cornell Univ., Ithaca, NY 14850), Kate Stafford (Univ. of Washington, Seattle, WA 98105), and Tina Yack (NOAA Southwest Fisheries Sci. Ctr., La Jolla, CA 92038)

Software analysis systems comprise an important stage in passive acoustic research. Here we compare and evaluate three animal acoustic analysis software systems: Ishmael, PAMGUARD, and XBAT. These packages are compared and evaluated for their capabilities at some of the common tasks in animal sound analysis: recording, display, detection, classification, measurement, localization, and tracking of animal sounds. They are also compared for their extensibility (how easy is it, say, to add a new detection algorithm), their ease of use (how hard is the software to learn to use, how quick to use it once one knows how), their hardware interfacing (what types of sound acquisition hardware can they receive sound from), their sound file interfaces (can they read everyone's sound files), their documentation, and their software interfaces (how well do they share data with other programs, other methodologies, like visual surveys). Special features of each system will also be discussed.

9:20

2aAB5. Passive acoustic monitoring of marine ecosystems in the Pacific Islands Region. Marc O. Lammers (Joint Inst. for Marine and Atmospheric Res., NOAA-CRED, 1125B Ala Moana Blvd., Honolulu, HI 96814, lammers@hawaii.edu), Kevin Wong, Russell E. Brainard (NOAA Fisheries, Honolulu, HI 96814), Whitlow W. L. Au (Univ. of Hawaii, Kailua, HI 96734), and Pollyanna Fisher-Pool (Joint Inst. for Marine and Atmos. Res., Honolulu, HI 96814)

Monitoring the changing state of marine habitats is a challenging task that is exacerbated when the habitats in question are in remote locations. Passive acoustic monitoring is sometimes the best, if not the only, means of gauging levels of biological and anthropogenic activities in such areas. Since 2006, the Pacific Islands Fisheries Science Center, in partnership with the University of Hawaii, has been building a network of long-term acoustic monitoring stations across the Pacific Islands Region using Ecological Acoustic Recorders (EARs). The network is currently composed of 29 long-term monitoring stations and spans the Hawaiian Archipelago, American Samoa, the Line Islands, Johnston Atoll, Wake Atoll, Guam, and the Commonwealth of the Northern Mariana Islands. EARs are diver-deployed and refurbished by local partners and/or during annual or biennial research cruises to deployment locations. A wide range of natural and anthropogenic acoustic signals are monitored, including sounds produced by invertebrates, fish, cetaceans, vessels, and surface weather events. The long-term trends in biological acoustic activity obtained through this network will be used to gauge the relative stability of the ecosystems associated with each location. Detections of vessels are common at many locations and provide a quantitative means of establishing levels of anthropogenic activity.

9:40

2aAB6. Habitat use of beluga whales revealed by acoustic loggers. Manuel Castellote (Res. Dept., L'Oceanografic of Valencia, Spain, mcastellote@oceanografic.org), Nick Tregenza (Chelonia Ltd., United Kingdom), Ruth Leeney (Provincetown Ctr. for Coastal Studies, MA), José Antonio Esteban (Res. Dept., L'Oceanografic of Valencia, Spain), Gregory O'Corry-Crowe (Harbor Branch Oceanograph. Inst., FL), Bill Lucey (Yakutat Salmon Board, AK), Kathleen M. Stafford (Univ. of Washington, Seattle, WA), Kit M. Kovacs, Christian L. Lydersen (Norwegian Polar Inst., Norway), Rauno Lauhakangas (Helsinki Inst. of Phys., Finland), Vera Krasnova, Anton Chernetsky, Alexandr Agafonov, and Vsevolod Bel'kovich (Russian Acad. of Sci., Russia)

The detection of odontocete echolocation signals with acoustic loggers has proved to be a very effective way to monitor their distribution, seasonal patterns and habitat use. However, belugas (*Delphinapterus leucas*) have never been monitored using this methodology. In order to evaluate its viability, echolocation signals of captive belugas were analyzed and the performance of T-POD (Chelonia Ltd., UK) acoustic loggers in captivity and in the Arctic environment was tested. Results from captivity demonstrated the potential of beluga remote monitoring by T-PODs. Field tests were organized in different environmental conditions with different populations of belugas to ensure the validity and detection robustness of this method. T-PODs were deployed in Svalbard (Norway) in Spring 2007, in Yakutat (Alaska) in Spring 2008, and in Solovetsky (Russia) in Summer 2008. Instruments were deployed concurrently with visual observations and hydrophone recordings to compare presence/absence data from the three different methods. Results proved that T-PODs positively detected all the identified periods of beluga presence without false detections and revealed temporal distribution correlations with tide levels, feeding behaviors, and circadian rhythms in the echolocation activity. Results from this project indicate that beluga remote monitoring is feasible in the Arctic environment using acoustic loggers.

10:00—10:20 Break

10:20

2aAB7. Passive acoustical counting of odontocetes using towed and stationed platforms. Tomonari Akamatsu (Natl. Res. Inst. of Fisheries Eng., Fisheries Res. Agency, 7620-7 Hasaki, Kamisu, Ibaraki 314-0408, Japan), Ding Wang, Kexiong Wang, Songhai Li, Shouyue Dong (The Chinese Acad. of Sci., Wuhan, 430072, China), and Satoko Kimura (Kyoto Univ., 606-8501 Kyoto, Japan)

Counting the number of animals is essential for monitoring and management of marine mammals. Autonomous stereo acoustic event recorder (A-tag) to identify each sound source enabled to count phonating odontocetes in the wild. Finless porpoises were surveyed by towed A-tags in 1100 km stretch of Yangtze River from Wuhan to Shanghai. The calculated detection probability using acoustic method was approximately twice that for visual detection. The difference in detection probabilities between the two methods was caused by the large number of single individuals that were missed by visual observers. However, acoustics tended to underestimate group size due to the limited resolution of sound source bearing angles. The detection performance regarding stationary acoustic monitoring of this species was also examined. Three stereo acoustic data loggers were placed at different locations near the confluence of Poyang Lake and the Yangtze River. Acoustic monitoring confirmed the low density bottle neck of the population between the lake and the river that was consistent with visual observations. The passive pulse event monitoring was proved to be a practical and useful alternative to visual observations, especially for long-term and range-wide monitoring.

Contributed Papers

10:40

2aAB8. Acoustic tracking of whistling dolphins offshore of Southern California. Kaitlin E. Frasier, Sean M. Wiggins, and John A. Hildebrand (Scripps Inst. of Oceanogr., Univ. of California San Diego, Ritter Hall 200E, La Jolla, CA 92093-0205)

Dolphin whistles were recorded near San Clemente Island, offshore of Southern California, by four autonomous high-frequency acoustic recording packages (HARPs) deployed in a sea-floor array (dimensions: 2 km $\times$ 2 km) centered around the R/P FLIP. This array layout allowed localization of the calling animals using time difference of arrivals (TDOAs) from the same calls simultaneously recorded on the four instruments. Whistle TDOAs were calculated using cross-correlations of spectrograms. A matched-field least-squares minimization procedure used whistle TDOAs to localize the animals. Successive localizations were combined to track movement over time, revealing local movements of presumably common dolphins (*Delphinus* spp.), typically during elevated nocturnal activity. These tracks, coupled with the animals' acoustic activity, provide insight into their behavior and could potentially be used to study their response to various stimuli including anthropogenic sources such as sonar. [This research is supported by CNO-N45.]

10:55

2aAB9. A real-time detection system for odontocete echolocation clicks in the low-energy processing environment of an acoustic glider. Holger Klinck (Cooperative Inst. for Marine Resour. Stud., and NOAA Pacific Marine Environ. Lab., Oregon State Univ. 2030 SE Marine Sci. Dr., Newport, OR 97365, Holger.Klinck@oregonstate.edu) and David K. Mellinger (Oregon State Univ., Newport, OR 97365)

This presentation will introduce a real-time system, the energy ratio mapping algorithm (ERMA), for detection of odontocete echolocation clicks. The system has been developed to run continuously over extended periods of time in low-energy processing environments such as acoustic gliders. The detector operates in the time domain and takes advantage of the species-specific differences in the power spectra and interclick interval of echolocation clicks to minimize the number of false positive detections. We will demonstrate the operation of ERMA using the clicks of beaked whales, family Ziphiidae, as example target species. To get information on the performance of ERMA, a comparison with more sophisticated and already established detection systems was conducted. Preliminary results indicate that the developed method is a promising tool to reliably detect echolocation clicks emitted by beaked whales in real time. On-going work is focused on implementation of the detector in the acoustic seaglider developed by the Applied Physics Laboratory of the University of Washington (APL-UW) and the use of this platform for autonomous passive acoustic monitoring and mitigation of beaked whales and other odontocete species.

11:10

2aAB10. Autonomous seafloor recorders and autodetection buoys to monitor whale activity for long-term and near-real-time applications. Christopher Tremblay, Tom Calupca, Christopher W. Clark, Matt Robbins, Eric Spaulding, Ann Warde (Bioacoust. Res. Prog., Cornell Univ., 159 Sapsucker Woods Rd., Ithaca, NY 148504, cjt35@cornell.edu), John Kemp, and Kristopher Newhall (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Biologists and engineers from Cornell have installed arrays of autonomous seafloor recorders in multiple ocean habitats. These systems are de-

signed to continuously monitor for the occurrence of endangered whales, particularly right whales; provide critical data on whale seasonal occurrence, distribution and relative numbers; and evaluate potential noise impacts from manmade activities and commercial shipping. Recent studies have been conducted in Massachusetts Bay around commercial LNG ports, the Arctic Ocean at seismic study locations, and in New York waters extending to the shelf edge. In combination with Cornell-developed MATLAB- and JAVA-based software applications, these seafloor recorders have been utilized to evaluate seasonal whale presence, calculate levels of commercial shipping and construction noise, and locate and track the movements and behaviors of individual whales under different acoustic conditions. Cornell and ocean engineers from the Woods Hole Oceanographic Institution have installed an operational network of autodetection buoys off New England to provide near-real-time data on right whale presence to transiting LNG vessels, government agencies, and public entities. When combined, the seafloor and autodetection buoy systems provide a powerful mechanism for understanding whale behavioral ecology and describing their acoustic habitats, while mitigating the risks from ship strike and noise exposure.

11:25

2aAB11. Passive acoustic survey of bowhead whales in the Chukchi Sea. Julien Delarue, Marjo Laurinolli, and Bruce Martin (JASCO Res., Ste. 432, 1496 Lower Water St., Halifax, NS B3J 1R9, Julien.Delarue@jasco.com)

Western Arctic bowhead whales (*Balaena mysticetus*) transit through the Chukchi Sea in spring and fall during their seasonal migration between their winter grounds in the Bering Sea and their summer feeding grounds in the Beaufort Sea, where their presence is well documented. Comparatively little is known about their use of the Chukchi Sea in summer and winter. Here results are presented from a long-term passive acoustic monitoring program using a multi-recorder array deployed from late July 2007 through July 2008 in this area. Bowhead calls were recorded intermittently between August 2007 and January 1, 2008 and again from May 27, 2008 until the end of the study. Of particular interest is the detection of songs in late fall–early winter, which had previously only been recorded during the northward spring migration. They were compared to songs recorded off West Greenland during the previous mating season and found to be different. This is the first evidence of geographic variation of bowhead songs. Overall these detections provide new information on bowheads' spatiotemporal use of the Chukchi Sea. The successes of the study also highlight the effectiveness of autonomous underwater acoustic recorders for marine mammal surveys in remote locations.

11:40

2aAB12. A high-resolution pressure tag for ocean depth measurements. Prajas John, Jaison Peter, Adrine Antony Correya, M. H. Supriya, and P. R. Saseendran Pillai (Dept. of Electron., Cochin Univ. of Sci. and Technol., Cochin- 682022, India)

Data Storage Tags form an important class of electronic tags, used for studying the migratory patterns of highly migratory marine species. In such tags, the parameters to estimate the position of the species are recorded in the tag memory and are retrieved subsequently upon recapturing the tag. The vertical and spatial migration of such marine species can be analyzed in detail using the pressure measurements made by the tags. The migratory behavior of marine species can be attributed to the light intensity variations and temperature profile as well as the search of food and suitable environmental conditions. The depth measurements can be made more accurate, if the resolution of the pressure sensor is significantly high. An absolute pressure measurement with a resolution of 1.2 mbar (approx. 1.2 cm in depth measurements) has been achieved with a pressure tag, capable of recording pressures at preset time intervals. Conversion of the hydrostatic pressure measurements into linear depth variation is carried out using the well-known equations. Pressure measurements were carried out using the pressure tag in the laboratory conditions as well as in the open ocean and the depth variations were computed using a user-friendly interactive software.

TUESDAY MORNING, 19 MAY 2009 BROADWAY I/II, 8:15 TO 11:20 A.M.

Session 2aAO

Acoustical Oceanography and Animal Bioacoustics: Acoustic Backscattering from Marine Life in the Ocean

Christopher D. Jones, Chair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105

Chair's Introduction—8:15

Contributed Papers

8:20

2aAO1. Theoretical models for imaging distributed fish groups in an ocean waveguide with wide-area sonar including multiple scattering. Mark Andrews and Purnima Ratilal (Northeastern Univ., 409 Dana Res. Ctr., Boston, MA 02115)

Wide area sonar has been shown to be a useful tool for instantaneously imaging shoals of fish distributed over large areas on the continental shelf. Numerical models are applied to simulate the active imaging system and determine the statistics of the received matched filtered intensity scattered off remote fish groups in the ocean waveguide environment. The model includes multiple scattering of the dense fish groups and uses a range-dependent acoustic model to simulate the ocean environment. The numerical model results are compared to analytic models, which make a single scattering assumption. We find the conditions for when multiple scattering is important and show that it is a function of both the fish density and mean target strength for a given distribution of fish. We use Monte Carlo simulations to explore the effects on the statistics of the received matched filtered intensity for an ocean waveguide randomized by internal waves. The model can be used to infer fish densities and abundances in wide area sonar imagery as well as provide error analysis of these estimates.

8:35

2aAO2. Atlantic herring low frequency target strength estimation from ocean acoustics wageguide remote sensing (OAWRS) data in the Gulf of Maine over 10 days of observation. Duong Duy Tran, Zheng Gong, Nikhila Srikanth, Mark Andrews, Purnima Ratilal (Dept. Elec. and Comput. Eng., Northeastern Univ., 360 Huntington Ave., Boston, MA 02115, tran.duo@neu.edu), and Nicholas Makris (MIT, Cambridge, MA 02139)

During the Gulf of Maine acoustic experiment in the fall of 2006, massive shoals of Atlantic herring were instantaneously imaged over wide areas using an ocean acoustics waveguide remote sensing (OAWRS) system during evening to midnight hours over a period of 10 days from Sept. 26 to Oct. 5. The low frequency target strength (TS) and abundances of the herring

population have been estimated by correlating the OAWRS data with localized measurements from a conventional fish finding sonar (CFFS) for 3 days from Oct. 1–3 [Gong *et al.*, J. Acoust. Soc. Am. **124**, 2586 (2008)]. Here we provide an analysis of the low frequency TS of the herring population in the frequency range from 300 to 1500 Hz and abundances for the remaining 7 days of the experiment. The acoustic scattering from herring populations is highly frequency dependent and is well modeled using a resonant scattering model for swimbladder bearing fish. Here we compare the TS estimates and the neutral buoyancy depth for herring over the observation period of 10 days.

8:50

2aAO3. Midfrequency backscatter imaging of fish schools in a shallow water waveguide. Christopher D. Jones and Darrell R. Jackson (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105, cjones@apl.washington.edu)

In this experimental effort, field data and modeling are used to investigate backscatter imaging of schools of pelagic fish in a shallow water oceanic waveguide. Two-dimensional (2-D) circular images of an oceanic waveguide are created using a monostatic circular receiving array and vertical line array source (12 kHz) deployed from a stationary ship. Horizontal multibeam images of a radial (depth integrated) cross section of the ocean are formed with targets and reverberation structure observed at ranges up to several Km. Examples of waveguide images are presented for data taken in the Puget Sound of Washington state. In some instances, the backscatter from aggregations of fish was observed to be higher than the reverberation; in other cases, fish schools could not be resolved. Fish school locations were independently observed by vertical echo sounding. Preliminary experimental results are shown to illustrate the potential of resolving the 2-D horizontal structure of aggregations of fish using such methods. Modeling is used to interpret the effects of waveguide propagation on imaging. Operational issues related to using waveguide imaging for fisheries' research and signal processing issues related to waveguide imaging will be discussed. [Work funded by NOPP with support from ONR and NOAA-OE.]

9:05

2aAO4. Tracking fish in a school using a broadband split-beam system. Masanori Ito, Ikuo Matsuo (Dept. of Information Sci., Tohoku-Gakuin Univ., 2-1-1 Tenjinsawa, Izumi-ku, Sendai, 981-3193, Japan), Tomohito Imaizumi, Tomonari Akamatsu (Fisheries Res. Agency, Kamisu, Ibaraki, 314-0408, Japan), and Yasushi Nishimori (Furuno Electric Co., Ltd, Nishinomiya, 662-8580, Japan)

Split-beam echo-sounder has been an important tool to understand fish behaviors undersea. In the split-beam system, each position of these fishes in three-dimensional space was estimated from delays of echoes and differences of delays between receive channels. To estimate the positions with high spatial resolution, the split-beam system which can emit and receive the broadband signals was constructed. However, in the case of a school of fish, it is difficult to determine accurately delay times of these fish because the echoes are overlapped with one another. To solve this problem, the new method was proposed to estimate delay times of fish by using the onset information of echoes, that is, extracting the steep temporal changes of the envelope pattern computed from the echo waveform. In a water tank experiment, it is examined that a location of a sphere can be accurately estimated by using the proposed method. In the sea, the echoes from schools of fish were measured with this system. It is examined that each location of fish in the three-dimensional space can be estimated and the numbers of fish can be counted. The tracks of the fish can be computed by changes of the fish locations along pings.

9:20

2aAO5. Broadband acoustic backscattering from live squid: Experiment and analysis. Wu-Jung Lee, Timothy K. Stanton, and Andone C. Lavery (Woods Hole Oceanogr. Inst., Dept. of Appl. Ocean Phys. and Eng., Woods Hole, MA 02543, wjlee@whoi.edu)

Squid are important organisms both ecologically and commercially. Acoustic scattering techniques can provide synoptic data on their distribution and abundance, and have the advantage of being efficient compared with traditional net sampling methods. However, knowledge of the scattering properties of squid is required to accurately convert the acoustic data into meaningful biological information. To better understand the scattering of sound by squid, and to facilitate future model development, a controlled laboratory backscattering experiment was conducted on live squid (*Loligo pealeii*) using broadband linear chirp signals (45–105 kHz) with data collected over the full 360 deg of orientation in the lateral plane, in 1-deg increments. The scattered spectra often showed significant structure over the frequency band available at different angles of orientation, in addition to high levels of ping-to-ping variability. Pulse compression signal processing techniques were also used, revealing the following dominant scattering features: (1) at normal incidence, the front and back interface of the animal were resolved and (2) at off-normal incidence, the anatomical features in the head region were found to dominate the scattering. This information can serve as the basis for an accurate acoustic scattering model for squid.

9:35

2aAO6. Humboldt squid distribution in three-dimensional space as measured by acoustics in the Gulf of California. Robyn S. Matteson, Kelly J. Benoit-Bird (College of Oceanogr. and Atmospheric Sci., 104 COAS Admin. Bldg., Oregon State Univ., Corvallis, OR 97331, rmatteson@coas.oregonstate.edu), and William F. Gilly (Stanford Univ., Pacific Grove, CA 93950)

The Humboldt squid, *Dosidicus gigas*, is a voracious migratory predator, important prey for many species, and the target of an economically important fishery. However, information on the behavior, life history, and biomass of this species is limited. Current knowledge is based almost exclusively on fisheries records, which only contain nighttime observations and provide no information on squid distribution and behavior outside of heavily fished areas. Acoustic measurements have long been used to study fish. Development of this method for application to squid provides a noninvasive technique to assess squid numbers and sizes. This method was applied in March and April of 2007 and November of 2008 in the Gulf of California, with a focus on the Santa Rosalia region. Squid densities were as high as 25 000 per km^3. Larger numbers of squid were observed at shelf breaks relative to other areas. Squid exhibited a marked diel vertical migration, with daytime depths of 300–400 m, bathymetry allowing, and nighttime depths of 0–200 m. In shallower coastal areas, the distribution of squid closely mirrored the shape of the bottom topography. This new technique is providing previously unattainable insights into the behavior and biology of this commercially and ecologically important squid species.

9:50

2aAO7. Acoustical and optical measurements provide evidence of ecological interactions in planktonic thin layers. Kelly J. Benoit-Bird and Timothy J. Cowles (College of Oceanic and Atmos. Sci., Oregon State Univ., 104 COAS Admin. Bldg., Corvallis, OR 97331, kbenoit@coas.oregonstate.edu)

Coincident physical, optical, and acoustical data were obtained in Monterey Bay, CA during an extensive study of planktonic layers. These concurrent data describing the distributions of phytoplankton, zooplankton, and small fish as well as gradients in the physical habitat were used to examine the interactions between vertically compressed plankton structures (thin layers), their consumers, and the local physical forces they experience. The steepness of the vertical gradient on the top versus the bottom of the plankton layer was correlated to the difference in the relative abundance of consumers above and below the layer. Phytoplankton layer gradients were steeper when more zooplankton were present on one side of the layer versus the other while zooplankton layers were more diffuse with a greater number of fish adjacent to one side of the layer than the other. Both layer types showed nearly symmetrical gradients when predators were in low abundance or absent. Differences in phytoplankton and zooplankton layer "shape" were not correlated with vertical gradients in shear or mixing potential surrounding layers. The consequences of trophic interactions within aggregations can be detected and assessed, and in the absence of strong physical gradients, grazers can play an important role in structuring plankton thin layers.

10:05—10:20 Break

10:20

2aAO8. Long-term broadband acoustic observations of zooplankton scattering layers in Saanich Inlet, British Columbia. Tetjana Ross (Dept. of Oceanogr., Dalhousie Univ., Halifax NS B3H 4J1, Canada; tetjana@dal.ca) and Gareth Lawson (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

The application of broadband techniques to fish and zooplankton bioacoustics is showing potential to transform the field into one that is much more quantitative. This is because broadband techniques allow the use of the known spectra of organisms or nonbiological sources of scattering to distinguish between scatterers, allowing discrimination without the need for extensive groundtruthing. This makes it ideal for remote monitoring of fish or zooplankton assemblages, since continuous net-sampling is often not possible. An upward-looking 85–155 kHz broadband sonar has been collecting data nearly continuously on the Victoria Experimental Network Under the Sea (VENUS) mooring in Saanich Inlet, British Columbia since March 2008. Saanich Inlet is known to have large populations of euphausiids, which create a strong acoustic scattering layer that migrates from depth to the surface and back each day. The thickness, timing, strength and spectral response of this layer is examined throughout the annual cycle and the feasibility of using this type of sonar to elucidate changing zooplankton assemblages (due to growth and changing species composition) is assessed.

10:35

2aAO9. Temporal patterns of fish and mesozooplankton near the Columbia River plume. Amanda M. Kaltenberg (College of Oceanic and Atmospheric Sci., Oregon State Univ., 104 COAS Admin. Bldg., Corvallis, OR 97331, akaltenb@coas.oregonstate.edu), Kelly J. Benoit-Bird (Oregon State Univ., Corvallis, OR 97331), and Robert L. Emmett (NOAA Northwest Fisheries Sci. Ctr., Newport, OR 97365)

The distribution of the dominant fish species in the California Current System is seasonal, with the timing and variability of these distributions having significant ecological consequences. The objective of this study was to observe the seasonal timing of fish school presence and to characterize the temporal patterns their abundance. Fish and mesozooplankton were observed near the Columbia River plume using 200-kHz bioacoustic moorings deployed throughout the transition period from winter conditions to summer upwelling conditions in 2008. Sea surface temperature, sea surface salinity, and ocean current data were used to characterize the oceanographic habitat. Acoustic mooring data revealed that the seasonal timing of schools of Pacific sardine and northern anchovy were linked with temperature and persisted after a rapid increase in temperature occurred near the end of May. The abundance of fish schools and mesozooplankton at two mooring stations was highly variable with dramatic fluctuations occurring over a period of hours to days that were linked with ocean conditions. The use of stationary acoustic moorings in this study provided an effective tool to monitor fish abundance at high temporal resolution and provides insight into the timing and variability of fish populations that will improve ecosystem-based management decisions.

10:50

2aAO10. Pilot work using a codend video camera for improved acoustic interpretation of backscatter observed during fisheries acoustic surveys. Lisa A. Bonacci and W. Waldo Wakefield (NOAA, Fisheries, NWFSC, 2032 SE OSU Dr., Newport, OR 97365, Lisa.Bonacci@noaa.gov)

One challenging but crucial element of fisheries acoustic surveys is accurate groundtruthing of the echo return. Assignment and proportioning of areas of backscatter to the correct species or mix of species affects fish biomass calculations. Typically, survey groundtruthing is done by midwater trawls targeting a single scattering layer of interest. However, the depth at which each species is caught is unknown. Therefore, this method is problematic when multiple scattering layers are present. Small scatters are also missed, which may be important contributors to the echo intensity. To address these issues, we conducted pilot work using a video camera mounted in the codend of an open midwater trawl, which enabled us to look at several scattering layers during each tow. For comparison, we conducted tows with a closed net and no camera at the same location. We then examined each acoustic echogram and the corresponding tow in order to compare backscatter calculations attributable to our species of interest. We found that there were some differences in these measurements between the two types of tows. We conclude that this method appears to provide improved knowledge of acoustic backscatter observed during fisheries acoustic surveys and plan to integrate this method into future surveys.

11:05

2aAO11. Density and soundspeed measurements of various Bering Sea zooplankton. Joseph D. Warren, Joy N. Smith (School of Marine and Atmospheric Sci., Stony Brook Univ., Southampton, NY 11901, joe.warren@stonybrook.edu), and Patrick H. Ressler (NOAA Fisheries AFSC, Seattle, WA 98115)

Walleye pollock is a commercially and ecologically important species in the Bering Sea. Assessment of pollock stocks is conducted regularly using a combination of trawl net sampling and multiple frequency active acoustic echosounder surveys. In addition to backscatter from the species of interest, other marine life including several types of zooplankton and nekton can contribute substantial amounts of acoustic backscatter energy which can provide information regarding the abundance and distribution of these animals as well. Acoustic scattering models are used to estimate both the amount and frequency response of the acoustic energy scattered by the zooplankton, but these models require knowledge of various properties that describe the acoustically important characteristics of the zooplankton. We collected live zooplankton from the Bering Sea during June and July 2008. The morphology (spatial dimensions) and material properties (density and sound speed) of individual and groups of zooplankton (and some nekton) were measured. The taxa studied included euphausiids (*Thysanoessa raschii, Thysanoessa inermis, Thysanoessa spinifera*), copepods (*Neocalanus sp.*), amphipods (*Themisto sp.*), gastropod (*Clione sp.*), siphonophores, jellyfish, and squid. The results of our measurements with regard to acoustic scattering model development and use will also be discussed.

Session 2aBB

Biomedical Ultrasound/Bioresponse to Vibration: Biomedical Applications of Acoustic Radiation Force

Mostafa Fatemi, Chair
Physiology and Biophysics, Mayo Clinic, Rochester, MN 55905

Chair's Introduction—8:00

Contributed Papers

8:05

2aBB1. Radiation force on spheres in helicoidal Bessel beams modeled using finite elements. David B. Thiessen, Likun Zhang, and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA, 99164-2814, thiessen@wsu.edu)

Analysis of the scattering of sound by spheres centered on ordinary and helicoidal (higher-order) Bessel beams [P. L. Marston, J. Acoust. Soc. Am. **124**, 2905–2910 (2008)] makes it possible to evaluate the radiation force on spheres centered on the beam in an ideal inviscid fluid. Potentially surprising results include cases where the radiation force is directed opposite the propagation direction of both types of beams [P. L. Marston (submitted for publication)]. For potential biomedical applications such as drug delivery, it would be necessary to know if a small transverse displacement of the sphere from the beam's axis causes a radiation force that pushes the sphere toward (or away from) the axis of the beam. Three-dimensional finite elements were applied to that problem. To trust FEM calculations of the radiation force with helicoidal beams, it was first necessary to verify that analytical values for the axial force are recovered in the on-axis helicoidal case since only the zero-order beam had been previously studied [D. B. Thiessen and P. L. Marston, J. Acoust. Soc. Am. **122**, 3025 (2007)]. Cases were found where displacement of a sphere from the beam axis causes a transverse restoring force. [Supported by ONR and NASA.]

8:20

2aBB2. Radiation torque on a sphere centered on an acoustic helicoidal (vortex) Bessel beam. Likun Zhang and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814, zhanglikun@wsu.edu, marston@wsu.edu)

Circularly polarized electromagnetic waves carry axial angular momentum and analysis shows that the axial radiation torque on an illuminated sphere is proportional to the power absorbed by the sphere [P. L. Marston and J. H. Crichton, Phys. Rev. A. **30**, 2508–2516 (1984)]. Hefner and Marston [J. Acoust. Soc. Am. **106**, 3313–3316 (1999)] proposed that a helicoidal acoustic beam (an acoustic vortex) also carries axial angular momentum and noted that absorption of such a beam should also produce an axial radiation torque. In the present work the acoustic radiation torque on a sphere centered on an acoustic helicoidal Bessel beam is analyzed and the torque is predicted to be proportional to the ratio of the power absorbed as a result of the interaction of sound with the sphere to the acoustic frequency. The torque is also proportional to the topological charge of the beam. Depending on the beam helicity, the torque is parallel or anti-parallel to the beam axis. The analysis uses a relation between the scattering and the partial wave coefficients for a sphere in a helicoidal Bessel beam [P. L. Marston, J. Acoust. Soc. Am. **124**, 2905–2910 (2008)]. [This research was supported in part by ONR and NASA.]

8:35

2aBB3. Acoustic radiation force on a gas bubble in tissue. Yurii A. Ilinskii, Evgenia A. Zabolotskaya, and Mark F. Hamilton (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713-8029)

The motion of a gas bubble subjected to acoustic excitation in a soft elastic medium such as tissue was analyzed previously assuming the radiation force acting on the bubble is the same as in liquid [Ilinskii *et al.*, J. Acoust. Soc. Am. **117**, 2338 (2005)]. In the present work we discuss corrections to the acoustic radiation force for finite values of the shear modulus. The analysis is based on the Piola-Kirchhoff equation in Lagrangian coordinates, in which only the stress tensor is nonlinear, and the equation is solved by perturbation. In the linear approximation an analytical solution is obtained for the scattered acoustic wave. The nonlinear stress and full radiation force is calculated at the next order of approximation. For negligible shear modulus the result for a liquid is recovered. For small but finite shear modulus the resulting force differs from that for a liquid by a factor that depends on $k_t R$, where R is bubble radius, $k_t = \omega/c_t$ the wavenumber, ω the angular frequency, and c_t the shear wave propagation speed. For $k_t R > 10$ the radiation force is practically the same as in liquid, but for $k_t R < 10$ its value can be significantly different. [Work supported by NIH DK070618.]

8:50

2aBB4. Optimization of ultrasound radiation force source for shear wave dispersion measurements. Matthew W. Urban and James F. Greenleaf (Dept. of Physio. and Biomedical Eng., Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905, urban.matthew@mayo.edu)

The objective of this study was to analyze and optimize the ultrasound radiation force source for creating shear waves for quantitative elasticity and viscosity measurements. A simulation model based on work by Bercoff *et al.*, [IEEE Trans. Ultrason. Ferroelectr. Freq. Contr. **51**(11), 1523–34 (2004)] was utilized to examine the shear wave propagation from a spatiotemporal impulsive force and the simulated radiation force generated from a linear array transducer. Parameters such as aperture size, F-number of the transducer, and medium shear elasticity and viscosity were varied to analyze how the shear wave amplitude and shape changes. The wave attenuation and group and phase velocities were evaluated. The aperture size and the F-number of the transducer affected the spatial distribution of the propagating shear wave significantly. The influence of the coupling term between the bulk and shear waves was analyzed spatially and temporally by evaluating parameters such as shear wave amplitude, wave attenuation, and the group and phase velocities. This simulation model provides insight for making shear wave propagation measurements from the perspective of optimizing the spatial and temporal characteristics of induced shear wave motion from an applied radiation force. [This work was supported in part by Grant EB002167 from NIH.]

9:05

2aBB5. Comparison of unconfined compression and spatially modulated ultrasound radiation force estimates of shear modulus. Stephen A. McAleavey, Erin Collins, Etana Elegbe, and Johanna Kelly (BME Dept., Univ. of Rochester, 309 Goergen Hall, Rochester, NY 14627, stephenm@bme.rochester.edu)

Spatially modulated ultrasound radiation force (SMURF) [S. McAleavey *et al.*, Ultrason. Imag. **29**, 87–104, (2007)] is a novel method for ultrasonic estimation of the low-frequency shear modulus properties of elastic media. In this approach, radiation force with a known lateral magnitude variation is

applied impulsively within a region of interest using short (~30 μs) bursts of ultrasound. The spatial frequency k of the radiation force intensity variation is selectable and determined through beamforming. Application of the radiation force impulse gives rise to a low-frequency (500–2000 Hz) shear wave. The temporal frequency ω of this wave is measured using Doppler ultrasound methods. The modulus is estimated from the relationship $G=\rho(\omega k)^2$, where ρ is the material density. To validate this method, the moduli of five samples of Zerdine (CIRS, Incorporated), an ultrasonically tissue-equivalent elastic material, were measured using conventional unconfined cyclic compression and SMURF implemented on a Siemens Antares scanner. The samples were cylindrical with diam 54 mm and height 25 mm. The shear modulus of the samples ranged from 2.5–35 kPa. For all samples, the two methods agreed to within the estimated variation. An intrasample variation of 8% was observed for SMURF in the (assumed uniform) samples. [Work supported by NIH/NIBIB.]

9:20

2aBB6. Optical tracking of acoustic radiation force impulse-induced dynamics in a tissue-mimicking phantom. Richard Bouchard (Dept. of Biomedical Eng., Duke Univ., Box 90281, Durham, NC 27708, rrb@duke.edu), Jason Streeter (UNC-CH and NCSU, Chapel Hill, NC 27599), Mark Palmeri, Gregg Trahey (Duke Univ., Durham, NC 27708), and Paul Dayton (UNC-CH and NCSU, Chapel Hill, NC 27599)

Optical tracking was utilized to investigate the acoustic radiation force impulse (ARFI)-induced response in a translucent, tissue-mimicking phantom. Suspended 10-μm microspheres were tracked axially and laterally at multiple locations throughout a microscope's field of view with 0.5-μm displacement resolution, in both dimensions, and at frame rates (with an attached video camera) of up to 36 kHz. Induced dynamics was successfully captured before, during, and after the ARFI excitation at depths of up to 4.8 mm from the phantom's proximal boundary. Results are presented for tracked axial and lateral displacements resulting from on-axis and off-axis (i.e., shear wave) acquisitions; these results are compared to matched finite element method modeling and independent, ultrasonically based empirical results. A shear wave reflection, generated by the proximal boundary, consistently produced an artifact in tracked displacement data later in time (i.e., after the initial ARFI-induced displacement peak). This new tracking method provides high-frame-rate, two-dimensional tracking data and thus could prove useful in the investigation of complex ARFI-induced dynamics in controlled experimental settings. [This work was supported in part by a National Science Foundation Graduate Research Fellowship.]

9:35

2aBB7. Acousto-magnetometry: A new vibrometery based on acoustic radiation force and magnetic measurement. Antonio Adilton O. Carneiro, Oswaldo Baffa (Departamento de Fisica e Matematica, Universidade de São Paulo, Ribeirao Preto, SP 14040-901 Brazil), Glauber T. Silva (Universidade Federal de Alagoas, Maceio, AL, 57072-970 Brazil), and Mostafa Fatemi (Mayo Clinic and College of Medicine, Rochester, MN 55905)

This work introduces a new method based on acoustic radiation force and magnetic measurement for interrogating the mechanical properties of fluid and biological tissues. In this method, ultrasound radiation is used to exert a low-frequency (in kHz range) force on a rigid magnetized target immersed in a viscoelastic medium. In response, the target vibrates in a pattern determined by viscoelastic properties of medium. We report the relation between the magnetic field signal and the incident ultrasonic pressure field in terms of the mechanical parameters of the medium. Simulations were conducted to demonstrate a simple approach based on using amplitude-modulated ultrasound to generate a dynamic acoustic radiation force on a magnetic target. The magnetic field generated by vibration of this target is then obtained and used to estimate the radiation-force-induced displacement as a function of time. It was observed that the intensity of the dynamic component of the magnetic field caused by the acoustic excitation is high enough to be registered by a conventional magnetic sensor. This proposed methodology presents a powerful tool for evaluation of acoustic radiation force as well as the mechanical properties of soft material.

9:50—10:20 Break

10:20

2aBB8. Shear wave dispersion ultrasound vibrometry of gel dosimetry phantoms. Silvio Vieira (Dept. de Fis. e Mat. da FFCLRP, Univ. de São Paulo, Av. Bandeirantes, 3900-Bairro Monte Alegre-Ribeirão Preto-SP, Ribeirão Preto, SP 14040-901, Brazil, silvio@pg.ffclrp.usp.br), Matthew Urban (Mayo Clinic, Rochester, MN 55901), Andre Baggio (Univ. de São Paulo, Ribeirão Preto, SP 14040-901, Brazil), Antonio Carneiro (Univ. de So Paulo, Ribeirão Preto, SP 14040-901, Brazil), and Mostafa Fatemi (Mayo Clinic, Rochester, MN 55901)

Shear wave dispersion ultrasound vibrometry (SDUV) has been introduced to use the dispersive nature of shear wave speed to locally estimate the material properties or tissue. Shear waves were created using a mechanical shaker and pulsed radiation force. Using the information from the phase one can estimate the shear velocity. A theoretical Voigt model allows us to find, at different frequencies (100–400 Hz), the viscosity and shear elasticity of the polymer gel dosimeter. The aim of this work was to perform the SDUV and mechanical vibrometry techniques to study the viscoelastic properties of absorbed dose phantoms. The experiments were performed using five phantoms of different absorbed doses varying 10 up to 50 Gy, while a nonirradiated phantom was kept as a control. The SDUV technique was performed using a pulsed radiation force to vibrate the gel provide by a confocal transducer with a focal depth of 70 mm, nominal frequency of 3 MHz. For the mechanical vibrometry and SDUV experiments, shear wave motion measurements were made with a commercial ultrasound machine with a research using a 10 MHz linear array transducer. The quantitative elasticity and viscosity results for these two experiments were compared.

10:35

2aBB9. Targeted ultrasound contrast agents—Mediated endothelial permeability. Pavlos Anastasiadis (Dept. of Mech. Eng., Univ. of Hawaii at Manoa, 2540 Dole St., Honolulu, HI 96822, pavlos@hawaii.edu), Chris N. Layman, and John S. Allen (Univ. of Hawaii at Manoa, Honolulu, HI 96822)

Endothelial cell dysfunction is considered to play a major role in the early stages of atherosclerosis. Targeted ultrasound contrast agents offer potential for therapeutic treatment though the mechanisms for enhanced ultrasound and microbubble mediated transport across endothelial layers are not well understood. In this study, we evaluate the endothelial cell-to-cell and cell-to-substrate gaps with the electric cell-substrate impedance sensing system (ECIS, Applied BioPhysics, Troy, NY). ECIS can detect the nanometer order changes of cell-to-cell and cell-to-substrate distances separately. Targeted ultrasound contrast agents (UTC) were employed in conjunction with the application of ultrasound, while continuous measurements of electrical resistance across the endothelial monolayers were conducted in real-time for the determination of permeability. The measurements were performed in both, the presence and absence of UCA. Additional complementary, noninvasive acoustic measurements of local changes in the mechanical properties of the endothelial monolayers were conducted at high-frequency (100 MHz) with a time-resolved scanning acoustic microscope. [This work was supported by the National Institutes of Health Grants NIH 2 P20 RR016453-05A1 and NIH 2 G12 RR0030161-21.]

10:50

2aBB10. Targeted ultrasound contrast agents for the imaging of biofilm infections. Pavlos Anastasiadis (Dept. of Mech. Eng., Univ. of Hawaii at Manoa, 2540 Dole St., Honolulu, HI 96822, pavlos@hawaii.edu), Kristina Mojica, Michelle L. Matter, and John S. Allen (Univ. of Hawaii at Manoa, Honolulu, HI 96822)

Targeted ultrasound contrast agents (UCA) offer a novel, potential method for diagnostic imaging of biofilm infections. Currently, there is no established method for molecular imaging of *in vitro* biofilm infections with any type of modality. For infective endocarditis, the case of biofilm formation on damaged or diseased heart valves, an early diagnostic method might greatly reduce the associated high mortality rate. Fluorescently labeled lectins were used with targeted ultrasound contrast agents in conjunction with epifluorescence microscopy and time-resolved scanning acoustic microscopy for visualization and characterization of the extracellular polymeric substances (EPS) of two types of infectious biofilms (*Staphylococcus aureus* and *Pseudomonas aeruginosa*). Noninvasive acoustic measurements of the biofilms were conducted at high-frequency (100 MHz) with a time-resolved scanning acoustic microscope. Measurements of the reflection and transmission coefficients, over a broad frequency spectrum for a range of biofilm growth stages in both, the presence and absence of targeted UCA were obtained. The biofilm morphology and structure were also investigated with acoustic microscopy as a function of growth. [This work was supported by the National Institutes of Health Grants NIH 2 P20 RR016453-05A1 and NIH 2 G12 RR0030161-21.]

11:05

2aBB11. Transcutaneous acoustic palpation (TAP) for the localization of peripheral, painful tissue. Pierre D. Mourad (Appl. Phys. Lab., Dept. of Neurosurgery, Box 356470, Univ. of Washington, Seattle, WA 98195, pierre@apl.washington.edu), Michel Kliot, and Jeffrey G. Jarvik (Univ. of Washington, Seattle, WA 98195)

Localizing painful tissue (pain generators) is central to medicine. Here we sought to demonstrate that intense focused ultrasound (iFU) can successfully differentiate normal from painful tissue. Rats were injected with complete Freunds adjuvant in a hind paw to create an inflammatory injury, or underwent a surgical procedure that damaged the nerve enervating the paw, creating a neuropathic injury. Each created a paw sensitive to external stimulation relative to the contralateral paw. iFU was applied individually to each hind paw in increasing doses until the animal withdrew either paw consistently from the iFU stimulus, thereby defining the iFU threshold dose for that animal. This data was correlated with paw withdrawal latencies to a heat lamp (Hargreaves) test. Sensitized paws responded to lower intensities and doses of iFU than control paws greater than 95 percent of the time with sensitivities and specificities generally greater than 90 percent. In general, iFU threshold tests and Hargreaves tests did not affect one another, a functional test of the safety of iFU. This preliminary evidence supports the hypothesis that iFU can safely discriminate between painful and normal tissue. Future work will use image-guided iFU to localize deep rather than superficial pain generators.

11:20

2aBB12. A stabilized forward elasticity finite element formulation yields a stable and convergent inverse elasticity solution. Carlos E. Rivas, Paul E. Barbone (Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, barbone@bu.edu), and Assad A. Oberai (Rensselaer Polytech. Inst., Troy, NY 12180)

"Elastography" refers to the procedure of using ultrasound to image tissue deformation. Images showing tissue deformation are sensitive to distributions of tissue mechanical properties. The mechanical properties themselves may be quantified by solving an inverse elasticity problem. An iterative approach to solving the inverse problem can be formulated by repeated solutions of the forward problem. That is, the shear modulus distribution sought is that which predicts a displacement field most consistent with the measured displacement field. Here we show that given plane displacement measurements for plane stress elasticity, the continuous elastic equilibrium equation uniquely determines the modulus distribution. On the other hand, we shall also demonstrate that the discrete elasticity equations from standard FEM discretization does not, even at the limit of infinite mesh refinement. We diagnose the problem as an underenforcement of the elasticity equations. With this knowledge, we have been able to design new forward elasticity FEM formulations that provide provably convergent forward *and inverse* elasticity solutions.

11:35

2aBB13. Ultrasonic properties of normal and cancerous human breast cell lines. Sang-Mo Shin (School of Medical System Eng., Gwangju Inst. of Sci. and Technol., 1 Oryong-dong, Buk-ku, Gwangju 500-712, Korea, sshin@gist.ac.kr), Koeng-Mo Sung, Hyun-Woo Koh (Seoul Natl. Univ., Seoul 151-600, Korea), and Pavlos Anastasiadis (Univ. of Hawaii at Manoa, Honolulu, HI 96822)

Time-resolved scanning acoustic microscopy (SAM) equipped with time-resolved signal analysis offers unique possibilities for studying the mechanical properties of cells *in vitro*. The method is nondestructive and noninvasive on a cellular level, which enables real-time measurements on living cells. The ultrasonic attenuation coefficients and sound speeds of MCF 10A normal breast cells and MCF 7 weakly invasive cancerous cells were measured and compared using a time-resolved SAM operating at 0.86 GHz center frequency and with 5 ns pulse duration. The attenuation coefficients of MCF 10A and MCF 7 cells were found to be 126 ± 67 and 443 ± 129 dB, respectively, while the sound speeds of the same cells were 1583 ± 24 and 1736 ± 61 m/s, respectively. Both the attenuation coefficient and sound speed of MCF 7 cells were higher than those of MCF 10A cells. In particular, the difference in the attenuation coefficients was pronounced. The relative differences in sound speed and attenuation coefficient could be utilized to differentiate normal cells from cancerous cells. [Weiss *et al.* IEEE Trans. Ultrasonics, Ferroeletrics, and Freq. Contrl. **54**, 2257–2271 (2007)] [Work supported by Korean Science and Engineering Foundation.]

Session 2aEA

Engineering Acoustics and Physical Acoustics: Lasers in Underwater Acoustics

Kenneth M. Walsh, Chair
K M Engineering Ltd., 51 Bayberry Ln., Middletown, RI 02842

Chair's Introduction—7:55

Invited Papers

8:00

2aEA1. Acquiring acoustic data using a scanning laser vibrometer. Walter H. Boober (Naval Undersea Warfare Ctr. Div. Newport, 1176 Howell St., Newport, RI 02841, walter.boober@navy.mil), Kim C. Benjamin, Stephen E. Forsythe (Naval Undersea Warfare Ctr., Newport, RI 02841), and Kenneth M. Walsh (K and M Eng., Middleton, RI 02842)

Scanning laser vibrometers have been in use for many years in acoustic measurements. Unique methods of characterizing both active and passive devices or structures with vibrational modes have been developed by NUWC personnel; these are combinations of signal generation techniques and physics-based data analysis tools. The signal generation techniques reduce the time for potentially lengthy scans at multiple frequencies through the use of a "comb" waveform, [J. Acoust. Soc. Am. **116**(5), (2004)], if the device under test is known to be linear. Scans can be done in either air or water. An analysis technique using the Helmholtz integral (using NUWC's MATLAB-based CHIEF code; Schenck and Benthein, NOSC 1970, 1988) allows analysis from an in-air scan of a transducer face (before encapsulation) to predict the transducer's far-field beam patterns under some conditions. This is especially valuable as a time-saving tool to ensure that the transducer is on track to achieve the final design intent. Examples of the techniques applied to various transducer designs will be presented; for comparison, single-tone results also will be shown.

8:25

2aEA2. On continuous-line acoustic array behavior in a laser Doppler sound pressure measurement system. Michael S. McBeth (Space and Naval Warfare Systems Ctr. Atlantic, 204 Indian Springs Rd., Williamsburg, VA 23185, m.s.mcbeth@ieee.org) and Yongxiang Hu (NASA Langley Res. Ctr., Hampton, VA 23681)

Continuous-line acoustic array behavior in a laser doppler anemometry (LDA) based sound pressure measurement system is considered. Traditional LDA ultrasound measurement systems intentionally minimize the laser beam intersection or interference fringe volume to better achieve point measurements of the sound field. The large laser beam diameters due to beam spreading in remote sensing applications from space or airborne platforms make this approach impractical. Since the sound pressure signal is integrated over the interference fringe volume, continuous-line acoustic array behavior is expected to be observed proportional to the length of the laser light path through the interference fringe volume. This behavior is analyzed from a theoretical perspective and an experimental setup to test for continuous-line acoustic array behavior in a LDA sound pressure measurement system is presented.

8:50

2aEA3. Measurement and imaging of high-frequency sonar fields using acousto-optic tomography. Pete Theobald, Stephen Robinson, Triantafillos Koukoulas, and Gary Hayman (Natl. Physical Lab., Hampton Rd., Teddington, Middlesex TW11 0LW, UK, pdt@npl.co.uk)

The acoustic near-field of a high-frequency sonar array transducer has been measured using a laser scanning acousto-optic method and reconstructed using a tomographic technique similar to that used for x-ray computed tomography. This measurement, using a laser interferometer to scan a plane in front of the transducer for a series of transducer rotations, provides an image of the acoustic field for a plane in front of the transducer array. This method allows measurement of the acoustic field without perturbing the field being measured, which can occur when using the planar hydrophone scanning method, provides high spatial resolution with minimal phase averaging across the measurement aperture and provides potential for significantly decreasing the scan time. Measurements have been performed on two 1–3 composite transducer arrays at frequencies between 330 and 500 kHz. These are compared with planar hydrophone scans obtained using a 1.5-mm probe hydrophone. Both parallel and fan beam tomographic scan geometries have been considered along with the effect of lateral scan resolution versus angular rotation resolution.

9:15

2aEA4. Laser light scattering by bubbles in water: Fundamentals and applications to acoustics. Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814, marston@wsu.edu)

Laser light scattering is sometimes used as an alternative to acoustical methods for monitoring bubbles in seawater. There has also been interest in using lasers to investigate bubbles in wakes. In some cases light scattering by bubbles has been used in conjunction with acoustical measurements to characterize dynamics of bubbles radiating sound [J. S. Stroud and P. L. Marston, J. Acoust. Soc. Am. **94**, 2788–2792 (1993)]. In applications such as these it is important to understand the optical properties of bubbles that differ significantly from drops and particles. Examples include critical angle scattering and the transition to total reflection [D. S. Langley and P. L. Mar-

ston, Appl. Opt. **23**, 1044–1054 (1984)], forward scattering and extinction [D. S. Langley and P. L. Marston, Appl. Opt. **30**, 3452–3458 (1991); J. S. Stroud and P. L. Marston, cited previously], glory back-scattering enhancements and shape effects [W. P. Arnott and P. L. Marston, J. Opt. Soc. Am. **A5**, 496–506 (1988); Appl. Opt. **30**, 3429–3442 (1991)] and Brewster angle scattering. Some optical effects of coatings on bubbles (which can occur naturally) have also been modeled [P. L. Marston, Appl. Opt. **30**, 3479–3484 (1991)]. [Research supported by ONR between 1980 and 1995.]

9:40

2aEA5. A laser-based method of detecting underwater sound through an ice layer. Lynn T. Antonelli and Fletcher A. Blackmon (Naval Undersea Warfare Ctr., Code 1512, 1176 Howell St., Newport, RI 02841, lynn.antonelli@navy.mil)

An aerial, remote sensor has been devised to detect underwater sound through an ice layer by optically measuring the resulting vibration velocity on the ice surface, using a laser Doppler vibrometer. The laser was focused on an ice block approximately 8 in. wide by 12 in. long and 3.5 in. thick floating in a tank of water 16 in. deep. A submerged transducer was used to generate underwater acoustic signals to provide ice surface vibrations for the sensor to detect. Although acoustic attenuation through the ice was not an issue for this test, laser alignment was crucial. Since sensor reliability depends on receiving laser reflections from the ice surface, the laser was steered nearly perpendicularly onto the ice surface that had both smooth and rough features. The broadband optical sensor provides a remote sensing capability for measuring ice surface vibrations in response to underwater acoustic signals such as from marine mammals, passive or active sonar, and from underwater acoustic communication systems. The results of this landmark experiment will be presented, which demonstrate the feasibility for detecting underwater sound through an ice layer by probing the ice surface with a laser vibrometer, paving the way for in-air sensing in arctic regions.

10:05—10:20 Break

10:20

2aEA6. Remote, aerial, translayer, nonlinear downlink underwater acoustic communication. Fletcher A. Blackmon (Naval Undersea Warfare Ctr., Code 1512, 1176 Howell St., Newport, RI 02841, fletcher.blackmon@navy.mil) and Lynn T. Antonelli (Naval Undersea Warfare Ctr., Newport, RI 02841)

The nonlinear mechanism for optical to acoustic energy conversion is explored for optoacoustic communication from an in-air platform to a submerged platform. This downlink communication can provide a bell ringer function and transmission of text and data. The nonlinear conversion mechanism, also known as the nonlinear optoacoustic regime where focused laser energy is converted to sound at the air-water interface, involves a phase change of the water medium through evaporation and vaporization leading to plasma production. The nonlinear conversion mechanism provides a more efficient, i.e., higher source level, controllable method for producing underwater acoustic signals. A means of deterministically controlling the spectrum of the underwater acoustic signal has been investigated and demonstrated by varying the laser-pulse repetition rate to provide *M*-ary frequency shift keyed signaling. This physics-based conversion process provides a methodology for providing low probability of intercept signals whose information is embedded in noise-like signals. These laser generated signals can then be used in a frequency hopped spread spectrum technique with the use of the proper receiver structures to take advantage of the frequency diversity and periodicity inherent in this type of signal structure that could also be used to combat frequency selective fading in underwater acoustic channels.

10:45

2aEA7. Intense underwater laser acoustic source for Navy applications. Theodore G. Jones, Melissa K. Hornstein, Antonio C. Ting (U.S. Naval Res. Lab., Plasma Phys. Div., 4555 Overlook Ave. SW, Washington, DC 20375, ted.jones@nrl.navy.mil), and Zachary W. Wilkes (Res. Support Instruments, Inc., Lanham, MD 20706)

An intense remote underwater laser acoustic source is under development at the Naval Research Laboratory. In a novel configuration, a tailored intense laser pulse can be designed to propagate many meters underwater and compress at a predetermined remote location. Controlled compression of these optical pulses is governed by a combination of optical group velocity dispersion and nonlinear Kerr self-focusing. Optical compression can result in laser-induced breakdown, localized heating, and acoustic shock generation. Recent experiments include near-field acoustic source characterization using lens-focused 400 and 800 nm pulses of a Ti:sapphire laser, as well as 532 and 1064 nm pulses of a YAG laser. Sound pressure levels over 210 dB were achieved using a compact laser. Acoustic source characterization includes measurements of photoacoustic energy conversion efficiency, acoustic power spectrum, and directivity. Nonlinear optical studies included the precise measurement of the Kerr index of water at 400 and 800 nm, as well as conditions for optical filament generation, and their effects on acoustic signals. Planned experiments include tests in a bubbly salt water tank. Experimental results will be presented, and laser sources and techniques for underwater acoustic generation will be compared. [This work is supported by the U.S. Office of Naval Research.]

11:10

2aEA8. Transducer characterization by laser Doppler vibrometry. Martin P. Cooling, Victor F. Humphrey (Inst. Sound & Vib. Res., Univ. of Southampton, Southampton, SO17 1BJ, UK, mc@isvr.soton.ac.uk), Pete D. Theobald, and Stephen P. Robinson (Natl. Physical Lab., Teddington, Middlesex, TW11 0LW, UK)

A laser Doppler vibrometer (LDV) can be set up such that its laser beam passes through the glass wall of a water tank to be incident upon the radiating surface of an immersed ultrasonic transducer. Scanning the laser beam across a transducer in this way has the potential to be a fast, noninvasive method for source characterization and, in turn, field prediction. Such measurements are, however, significantly complicated by the acousto-optic interaction—that is, the effect on the measurements of the acoustic field through which the laser beam passes. To increase understanding of the resultant acousto-optic artifact, a detailed simulation of the LDV measurement of a circular, plane-piston transducer emitting a tone-burst has been created. The use of a transient field is important for simulation and experiment, such that measurements can be made over a time window which ends before any acoustic signal reaches the water tank boundaries. The simulation results show a significant acousto-optic artifact contribution to the surface velocity data, but also that for some applications useful field predictions may be made in spite of this. To complement the simulations, experimental measurements have been made using a commercial LDV (Polytec PSV-400) on a 500 kHz circular transducer.

11:25

2aEA9. Measurement of the resonance frequency of a bubble using a laser Doppler vibrometer. Theodore F. Argo, IV and Preston S. Wilson (Mech. Eng. Dept. and Appl. Resource Labs., The Univ. of Texas at Austin, Austin, TX 78712-0292)

Knowledge of the behavior of bubbles that are confined within tubes or channels has a direct impact on the development of various medical applications and industrial processes. In such confined spaces, bubbles are not easily observed by traditional experimental means, such as acoustically with hydrophones, or optically with Mie scattering or stroboscopy. Further, facilitating observation with these methods may significantly perturb the system. A laser Doppler vibrometer (LDV) requires only a narrow (~1 mm diam) line-of-sight access for the beam and illumination of the bubble does not perturb its dynamics. LDV measurements of the resonance frequency and quality factor of an air bubble suspended in a small water tank are presented to demonstrate the utility of this technique. The measurements are compared to a model that includes the acoustic effect of the tank walls. The precision of the technique was found to be similar to the precision of a traditional hydrophone-based acoustic technique.

TUESDAY MORNING, 19 MAY 2009 EXECUTIVE SALON II/III, 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 State Univ., Dept. of Chemistry and Physics, Terre Haute, IN 47809

Chair's Introduction—10:00

Approximately 20 acoustics experiments will be set up, ranging in complexity from simple resonance on a string to ultrasonic levitation. Around 40 local high school science students will perform the experiments with help from ASA scientists and students. Regular ASA conference participants are welcome to the session as long as they do not interfere with student experimentation.

Session 2aMU

Musical Acoustics: Wind Instruments I

Thomas D. Rossing, Chair
Stanford Univ., CCRMA, Stanford, CA 94305

Invited Papers

8:00

2aMU1. Three decades of wind instrument research at Brigham Young Univ. William Strong (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

Wind instrument research at Brigham Young University during the 1970s, 1980s, and 1990s will be reviewed. In more or less chronological order, the research was as follows: (1) Calculation of input impedances of an oboe; (2) predictor coefficient analysis /synthesis of French horn tones; (3) functional model of a simplified clarinet; (4) calculated and measured modal frequencies of brass instruments; (5) calculation and measurement of flute impedances and standing waves; (6) calculation and measurement of horn impedances; (7) simulation of a player-clarinet system; (8) optimization of clarinet toneholes; (9) simulation of a player-trumpet system; (10) stroboscopic measurement of the lip motions of a trombone player; (11) calculation and measurement of nonlinear effects in loud trombone tones; (12) finite element modeling of the lips of a trombone player. [Work supported by the Institute for Research in Musical Acoustics.]

8:30

2aMU2. Spectral component relationships in wind instrument transients. Peter L. Hoekje (Dept. of Phys. and Astron., Baldwin-Wallace College, 275 Eastland Rd., Berea, OH 44017, phoekje@bw.edu)

Playing characteristics of a wind musical instrument are related to its transient response, the efficiency of its sound production, and the stability of its sustained tones, as determined partly by the acoustical resonances of the instrument, and their nearly harmonic relationships. These can be explored using analysis of instrument attack transients by tracking development of the spectral components. A coupled oscillator model is used to explain the benefits to an instrument's response due to resonances at harmonics of the playing frequency. Furthermore, the slight deviation from harmonic relationships of an instrument's resonances leads to a shift in its tone center during a crescendo. A resonance at a frequency above the tone-hole lattice cutoff can affect the playing behavior of flutes and cornettos. The player's airway in reed woodwinds and brasses is as much a part of the acoustic system as the instrument proper, and its resonances can have a correspondingly important effect.

Contributed Papers

9:00

2aMU3. Impulse response reverb as a tool for voicing organs. Michael Steppat (Erlenweg 17, D-14558 Nuthetal OT Saarmund)

The sound of a flue pipe can be adjusted by voicing. With parameters like the cut-up of the upper lip and their width, the opening in the foot, the height of the languid, and the width of the slit the voicer can influence the promptness of speech and adjust the intensity of the starting transient. The acoustical environment (especially the reverberation time) plays an important role in the reception of the starting transient, therefore the organ builder has to also take the architecture into account. Convolution reverb allows a simulation of the reverb at different positions. Sampled organ sounds with modified starting transients or modeled sounds can be played on a computer and mixed with a convolution reverb using measured impulse responses from different positions of the room. This computer model can be used during the design process to estimate the parameters for the voicing. Results of measurements in a church building and different modeling techniques will be presented

9:15

2aMU4. Measuring the Young's modulus of reeds from American reed organs. Michael B. Wilson (Juniata College, 1700 Moore St., Huntingdon, PA 16652, wilsomb05@juniata.edu), Alexander M. Ramm, and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

A method of wave coefficients has been used to measure the Young's modulus of brass reeds from American reed organs. Although for some purposes these reeds can be approximated as uniform cantilever beams, their cross sections are sufficiently nonuniform that simple methods such as measuring beam displacement for a known applied force or using resonant frequencies of transverse and torsional modes of vibration are not sufficient to obtain accurate results. In the wave coefficient method [Y. Liao and V. Wells, J. Sound Vibr. **295**, 165–193 (2006)] a wave number is found that forces data at all measurement points to conform to the general forced-vibration solution for a beam, implying that they have the same wave coefficients. This method also assumes a uniform cross section but does not require measurements along the entire length of the beam and was applied only to the most uniform sections of the reeds. Three reeds of the same frequency were selected, with cross sections of varying uniformity. The resulting measurements of the Young's modulus seemed consistent with expected values for brass. For one reed with nearly uniform cross section the wave coefficient method was in good agreement with the simpler methods.

9:30

2aMU5. The influence of curvature on the vibration and tone quality of pipe organ reed tongues. G. R. Plitnik (Dept. of Phys., Frostburg State Univ., Frostburg, MD 21532) and J. Angster (Fraunhofer Inst. fur Bauphysik, Stuttgart, Germany)

Given certain design constraints, such as the type of stop being voiced and the desired tone quality, organ reed voicers must curve each tongue so as to produce the best and most stable tone, as well as maintaining a consistent tone quality across an entire rank of pipes. This study investigated the effects of varying tongue curvatures on two pipes, an F2 Trompette and an

F2 Clarinet. Reed vibration was measured under typical conditions by laser vibrometer; the pressure waves in the boot and shallot were measured by means of one-eighth inch microphones and the emitted sound was recorded at the egress. By integrating the velocity data, the displacement of the reed may be plotted in tandem with the pressure difference across the reed, allowing tongue position to be plotted versus the pressure variation during one steady-state cycle for each reed while vibrating. These graphs can then be compared and correlated to the observed tone quality for each of the differently curved reeds.

9:45

2aMU6. Acoustics of single-reed duck calls. Nicholas P. Goodweiler (College of Eng., Univ. of Iowa, Iowa City, IA 52242, nicholas-goodweiler@uiowa.edu) and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

A single-reed duck call features a reed of Mylar or similar elastic material mounted above a curved tone channel. The arrangement resembles that of a clarinet, except that the extreme curve of the tone channel (compared to the relatively flat lay of the clarinet) allows for a rather large volume airflow. In addition, the flexibility of the reed results in a low fundamental frequency (typically 100–140 Hz) so that the pitch of the call is close to the natural frequency of the reed, modified only slightly by the short resonating barrel. As the blowing pressure is increased, there is a gradual rise in pitch, punctuated by periodic jumps from one apparent regime of vibration to another. As the blowing pressure is lowered, some hysteresis effects are observed. Several calls built by different makers have been studied, each with a differently proportioned barrel and tone channel. Some of the phenomena observed include reed velocity and displacement, changes in effective reed length as a function of blowing pressure, and effects of barrel design on tone quality. Some observations have also been made of the volume airflow and effects of attaching pipe resonators.

TUESDAY MORNING, 19 MAY 2009

MAIN AUDITORIUM,
PORTLAND BUILDING, 8:00 A.M. TO 5:00 P.M.

Session 2aNS

Urban Design with Soundscape in Mind

A Symposium on Urban Planning with the Consideration of Noise Impacts and the People Concerned

8:15

Introduction to Symposium and Keynote Speaker
Kerrie Standlee. P.E., Daly-Standlee & Associates, Inc., Beaverton, OR

8:20

The Political Need to Consider Noise Impacts on Urban Livability
City of Portland Government Official

8:45

The Physical Need to Consider Noise Impacts on Urban Livability
Kerrie Standlee. P.E., Daly-Standlee & Associates, Inc., Beaverton, OR

9:10

The Historic Approach in Addressing Noise Impacts on Urban Livability

9:15

Noise Ordinances and Zoning Laws
Paul Herman, Retired, Noise Control Officer, City of Portland, Portland, OR
Paul Van Orden, City of Portland Noise Control Officer, Portland, OR
Jerry Wilson, Retired, Oregon DEQ Environmental Noise Anaylst
Ioana Park, P.E., BRC Acoustics, Seattle, WA

10:15 — 10:30 Break

10:30

Use of Building Site and Design to Address Noise Impacts
Bennett Brooks, P.E., Brooks Acoustics Corporation, Vernon, CT
Daniel C. Bruck, Ph.D., BRC Acoustics, Seattle, WA

11:15

Use of Noise Mapping to Address Noise Impacts
Ken Kaliski, Resources Systems Group, Inc., White River Junction, VT

12:10 — 1:30 Lunch

1:30

A New Approach in Addressing Noise Impacts in an Urban Setting-The need to consider people's minds

1:35

Soundscape in the Urban Environment — What does it mean
Brigitte Schulte-Fortkamp, Technische Univ. Berlin, Inst. of Fluid Mechanics and Engineering, Berlin, Germany
Bennett Brooks, P.E., Brooks Acoustics Corporation, Vernon, CT

2:35

Application of Soundscape Techniques
Gary Siebein, School of Architecture, University of Florida, Gainesville, FL

3:35

Closing Remarks
Kerrie Standlee. P.E., Daly-Standlee & Associates, Inc., Beaverton, OR

3:50

Tour of Portland Eastside Esplanade for Hands-On Demonstration of Soundscape Techniques
Tour led by Soundscape Presenters

5:00

Symposium Adjourned

TUESDAY MORNING, 19 MAY 2009 BROADWAY III/IV, 8:00 TO 10:15 A.M.

Session 2aPAa

Physical Acoustics: Cavitation: Mechanisms and Physical Effects

Thomas J. Matula, Chair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105-6698

Contributed Papers

8:00

2aPAa1. Laser-induced dielectric breakdown as a method for noncontact pressure measurement. Jonathan R. Sukovich, Todd W. Murray, and R. Glynn Holt (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Being able to accurately measure pressure fields in acoustic experimentation is of critical importance for understanding and describing observed acoustical phenomena. However, in many experimental situations, it is undesirable or impossible to measure the acoustic field using a transducer in contact with the acoustic host medium. It becomes useful then to have alternative, noncontact, methods for reliably and accurately measuring pressure fields in experiments. One alternative to traditionally invasive means, such as the use of a hydrophone, to measure pressure, is to use focused, short (5–7 ns), high power laser pulses to cause dielectric breakdown. Breakdown events can be detected by looking for the signature, near instantaneous, flash of broadband light generated by the subsequent ionization cascade. This procedure can then be repeated under different known hydrostatic pressures to establish pressure's effect on the breakdown strength of water. Once a relationship is established, the process can be used to map local pressures in acoustic fields of sufficiently long period as to be considered hydrostatic in relation to the laser pulse duration. [Work supported by the US Army Space and Missile Command.]

8:15

2aPAa2. Acoustic measurement of bubble size and position in an ink jet printhead. Arjan van der Bos, Roger Jeurissen, Michel Versluis, Detlef Lohse (Phys. of Fluids, Dept. of Appl. Phys., Univ. of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands), Hans Reinten, Herman Wijshoff, Marc van den Berg, and Jos de Jong (Oce Technologies B.V., 5900 MA, Venlo, The Netherlands)

A bubble can be entrained in the ink channel of a piezo-driven ink jet printhead leading to malfunction. Here, an acoustic sizing method for the volume and position of the bubble is presented. The influence of the bubble on the channel acoustics is detected by the piezo-actuator operating in a sensing mode. The acoustic response is also modeled in a linearized bubble dynamics model and it is demonstrated how the size of the bubble and its position within the channel can be inferred from the inverse problem. The validity of the presented method is supported by time-resolved optical observations of the bubble dynamics visualized in a glass connection channel. The combined optical and acoustical recordings are compared to the model, where a very good agreement is found. [This work is part of the research program of the Technology Foundation STW and financially supported by Océ Technologies B.V.]

8:30

2aPAa3. Laser nucleation of bubble clusters in a spherical resonator. Phillip A. Anderson, Todd W. Murray, and R. Glynn Holt (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Due to the strong energy focusing found in sonoluminescence, there is much interest in high-energy bubble collapses. There is evidence in the literature that the collective collapse of a bubble cluster is more violent than a single bubble collapse under similar conditions. Precise control over the spatial and temporal placement of bubbles in an acoustic field would seem critical for achieving the strongest collapses, but nucleating bubbles acoustically is not precise or repeatable and does not allow for the creation of arrays of bubbles. A method is presented for creating two- and three-dimensional bubble arrays in a controllable way. A pulsed Nd:YAG laser is used with custom diffraction gratings and focusing optics to nucleate bubble arrays in a spherical resonator at high static pressure. A specific implementation will be presented, along with results addressing uniformity and repeatability. [Work supported by the US Army Space and Missile Command.]

8:45

2aPAa4. Collapse dynamics of laser-nucleated bubble clusters in a spherical resonator. Phillip A. Anderson, Todd W. Murray, and R. Glynn Holt (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

The collective collapse of bubble clouds or clusters is important to many applications, including underwater sound propagation, shock-wave lithotripsy, and focused ultrasound therapy. Energetic collective cluster collapses appear to happen only when there is sufficient nuclei density. In order to study this phenomenon, we have implemented a laser-based method for creating clusters consisting of planar or cubic arrays of bubbles with controllable interbubble spacing. The growth and collapse of the arrays are observed using ultrahigh-speed photography and other diagnostics. Effects of interbubble spacing and nucleation phase with respect to the acoustic period are studied. [Work supported by the US Army Space and Missile Command.]

9:00

2aPAa5. Detecting cavitation in liquid mercury exposed to a high-energy pulsed proton beam. Nicholas J. Manzi, Robin O. Cleveland, R. Glynn Holt, and Ronald A. Roy (Dept. of Mech. Eng., Boston Univ., Boston, MA 02115)

The Oak Ridge National Laboratory spallation neutron source (SNS) employs a high-energy proton beam incident on a liquid mercury target to generate short bursts of neutrons. Concomitant with neutron production is the rapid heating of the mercury, resulting in the formation of acoustic shock waves and the nucleation of cavitation bubbles. The subsequent collapse of these cavitation bubbles may result in the eventual erosion of the target's steel walls. We report on preliminary measurements using two passive cavitation detectors (megahertz frequency focused and unfocused PZT transducers), installed in a mercury test target to monitor cavitation activity for proton beam energies ranging from 0.041 μC to 4.1 μC. Cavitation was initially detected for a beam "charge" of 0.205 μC by the presence of an acoustic emission approximately 120 μs after arrival of the incident beam, which is characteristic of inertial cavitation collapse. As beam energy was increased, the collapse time increased until the lifetime of the bubbles exceeded the reverberation time of the chamber (~300 μs) at which point complex emissions signatures were detected due to the interaction of the reverberation signal and the cavitation bubbles.

9:15

2aPAa6. Production of cavitation erosion damage by means of an electromagnetic shock wave source. Qi Wang, Nicholas J. Manzi, R. Glynn Holt, Ronald A. Roy, and Robin O. Cleveland (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, robinc@bu.edu)

Central to a spallation neutron source (SNS) is a vessel of liquid mercury. During SNS operation acoustic pulses, with pressures in excess of 10 MPa, are generated in the mercury likely resulting in cavitation, which may erode the walls. We developed an electromagnetic shock wave source to generate similar pressures and cavitation. An electromagnetic source consists of a coil placed close to a metal plate, which is in contact with a fluid. The source is driven by discharging a capacitor through the coil, which results in a repulsive force on the plate launching acoustic waves in the fluid. A 75 mm diameter source coupled to water was fabricated, and the plate motion was measured by laser Doppler vibrometry and the acoustic field by a piezoelectric needle for different capacitance, charging voltage, number of coils, and plate properties. For a 2 μF capacitor charged to 12 kV and discharged through a 28-turn coil with a 0.8 mm thick aluminum plate (spherically focused) the peak pressure measured in water was 20 MPa. Cavitation clouds were produced over a wide range of voltages (imaged with 15 μs exposure) with lifetimes around 300 μs and produced erosion craters on a porous solid. [Work supported by ORNL-SNS, UT-Battelle, US-DOE Contract DE-AC05-00OR22725.]

9:30

2aPAa7. The ultreo: A novel toothbrush that combines sonic and ultrasound physics. Pierre D. Mourad (Appl. Phys. Lab. and Dept. of Neurosurgery, Box 356470, Univ. of Washington, Seattle, WA 98195, pierre@apl.washington.edu), Frank Roberts (Dept. of Peridontics, Univ. of Washington, Seattle, WA 98195), and Chris McInnes (Ultreo, Inc.)

Efficacious, daily oral health care represents an important part of maintaining the overall health of individuals. Here we sought to develop and demonstrate the usefulness of a new power toothbrush, one that incorporated both sonic processes (in the form of rapid bristle motion that directly removes plaque as well as generates bubbles) and ultrasound (sourced from the brush head, that activated the bubbles). We adapted a set of existing power toothbrushes to incorporate an ultrasound source with sufficient controls to explore the effects of different ultrasound parameters on plaque removal *in vitro* and *in vivo*. The combination of sonic and ultrasound physics removed plaque *in vitro* in a synergistic manner, likely through the action of cavitation. *In vivo*, the Ultreo toothbrush—a commercial version of this brush—removed comparable amounts of plaque to another commercial brush in half the time, and removed more plaque than manual toothbrushes. A combination of sonic bristle motion and ultrasound-facilitated cavitation appears to produce superior plaque removal compared to existing, sonic only, or manual tooth brushes.

9:45

2aPAa8. Motion of dissolved gas in a cavitating fluid. I. V. Mastikhin, B. Newling, and S. Kristoffersen (Dept. of Phys., Univ. of New Brunswick, 8 Bailey Dr., Fredericton, New Brunswick, E3B 5A3, Canada)

A strong acoustic field in a liquid separates the liquid and dissolved gases by the formation of bubbles (cavitation). Bubble growth and collapse is the result of active exchange of gas and vapor through the bubble walls with the surrounding liquid. This work investigates a new approach to the study of cavitation, not as an evolution of discrete bubbles, but as the dynamics of molecules constituting both the bubbles and the fluid. Direct Magnetic Resonance Imaging measurements of dynamics of the liquid and the dissolved gas show that the motions of dissolved gas (freon-22, CHClF2) and liquid (water) can be very different during acoustic cavitation and are strongly affected by the ultrasonic power, filtration, or previous cavitation of the solvent. The observations suggest that bubbles can completely refresh their content within two acoustic cycles and that long-lived (~minutes) mi-

crobubbles act as nucleation sites for cavitation. The technique is complementary to the traditional optical and acoustical techniques. [V. Mastikhin and B. Newling, Physical Review E (2005, 2008).] [Authors thank the Natural Science and Engineering Research Council of Canada for funding.]

10:00

2aPAa9. Simultaneous, noninvasive determination of fluid viscosity and density from acoustical resonances. Curtis F. Osterhoudt and Dipen N. Sinha (Los Alamos Natl. Lab., MPA-11 MS D429, Los Alamos, NM 87544, sinha@lanl.gov)

U-tube densitometers make fluid density measurements based on eigenfrequency shifts, which depend on the fluids contained within the densitometer. This measurement method is the current standard for density measurements in industrialized settings. However, ruggedized densitometers have several drawbacks. They are expensive, and require diverting fluids into the apparatus to make their measurements. Viscometers, too, are considered invasive, or require diverting fluid into a measurement chamber (the one exception is the "OSCAR" viscometer, which cannot make measurements on flowing systems). In contrast, a simultaneous density/viscosity measurement can be made without diverting fluid from the container or pipe carrying it. The measurement, in which the container (e.g., pipe section) itself becomes the sensor, is based on acoustical frequency resonance shifts and resonance width changes. Such measurements can in principle be made at a standoff from the container, reducing price and requiring no further contact with the fluids. This presentation will address the technique, give the theoretical background for it, and present experimental validation.

TUESDAY MORNING, 19 MAY 2009 BROADWAY III/IV, 10:30 A.M. TO 12:00 NOON

Session 2aPAb

Physical Acoustics: Thermoacoustics

Martin D. Verweij, Chair
Lab. for Electromagnetics Research, Delft Univ. of Technology, 2628 CD Delft, The Netherlands

Contributed Papers

10:30

2aPAb1. Thermoacoustics in irregular media. Carl Jensen and Richard Raspet (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

One approach to acoustics in random porous media is to use fitting parameters to match theoretical models derived from exact parallel pore theory to measured values of the media's acoustic properties such as the complex compressibility and complex density. [Roh *et al.* J. Acoust. Soc. Am., **121**, 1413–1422 (2007)] extended these ideas to the more complicated thermoacoustic case of an applied temperature gradient by using empirical factors to reshape the known solution for thermoacoustics in a regular framework of rigid tubes to RVC and aluminum foam samples. The validity of this model has been investigated by directly simulating the sound propagation through several irregular, 3-D geometries using a thermal fluid solver. The acoustic and thermoacoustic simulation results for several geometries will be presented and compared to the theoretical results.

10:45

2aPAb2. Helmholtz-like resonators for thermoacoustic prime movers. Bonnie J. Andersen (Dept. of Phys., Utah Valley Univ., 500 W. University Pkwy., Orem, UT 84058) and Orest G. Symko (Univ. of Utah, Salt Lake City, UT 84112)

In a thermoacoustic prime mover, high acoustic output power can be achieved with a large diameter stack and with a cavity with a large volume attached at the open end of the resonator containing the stack. The combination of resonator and cavity makes the device Helmholtz-like, with special characteristics of the resonant frequencies and quality factor Q. Analysis of its acoustic behavior based on a model of a closed bottle presents features that are useful for the development of such prime movers for energy conversion from heat to sound. In particular, the arrangement produces in the cavity a high sound level, which is determined by the Q of the system. Comparison with a half-wave resonator type of prime mover, closed at both ends, shows the advantages of the Helmholtz-like device.

11:00

2aPAb3. Ultrasonic thermoacoustic prime movers. Myra Flitcroft and Orest G. Symko (Dept. of Phys., Univ. of Utah, 115 S. 1400 E., Salt Lake City, UT 84112)

Thermoacoustic devices that generate sound with frequencies in the near ultrasonic range (~21 kHz) have been developed. These devices consist of a wave resonator that is 4.1 mm in length, with hot and cold heat exchangers made of copper mesh placed at the resonator's midpoint. Between these is sandwiched a stainless steel stack; the working fluid is air at 1 atmosphere. Heat is injected to the hot heat exchanger, creating a temperature gradient across the stack. When this exceeds a critical temperature gradient, acoustic oscillations ensue with intensities over 140 dB. The performance of a given engine depends drastically on the alignment achieved in the assembly process. Due to this variability, temperature differences at onset ranged from 85°C to 180°C. In order to increase positive feedback in the systems, single engines were coupled to a resonant cavity, the end of which was fitted with a pressure transducer. The use of a cavity also provides an acoustic coupling mechanism in which two ultrasonic engines can be attached to the same cavity, allowing for the investigation of synchronization of these devices in this regime. [Research supported by the U.S. Army Space and Missile Defense Command/U.S. Army Forces Strategic Command.]

11:15

2aPAb4. Miniature traveling wave thermoacoustic engine. Ivan A. Rodriguez and Orest G. Symko (Dept. of Phys., Univ. of Utah, 115 S. 1400 E., Salt Lake City, UT 84112)

Five high frequency annular thermoacoustic traveling wave devices have been developed and characterized. A 1.27 cm bore diam 2 kHz engine was optimized, increasing its maximum acoustic output from 140 to 167 dB. A new 1.60 cm bore diam 2 kHz engine has been assembled with the expectation of even higher acoustical power output because power is proportional to cross-sectional area. In the 3 kHz regime, two geometries were investigated, an oval shape and more square version, which is easier to assemble. These devices were found to have comparable outputs suggesting that the square geometry is a viable alternative, and, thus, all additional devices were constructed this way. A 4 kHz engine was also studied. In order to measure the effects of the heat exchangers and regenerator assembly on the acoustic flow, a resonator tube was constructed. This tube was fitted with the device innards and driven by a speaker. An absorption coefficient of 0.4 was measured. This information along with the 916 T across the regenerator allows for an estimate of the limit on the acoustic gain of the system to be

calculated. [Research supported by the U.S. Army Space and Missile Defense Command/U.S. Army Forces Strategic Command.]

11:30

2aPAb5. Particle image velocimetry study of acoustic field in miniature traveling wave device. Ivan A. Rodriguez and Orest G. Symko (Dept. of Phys., Univ. of Utah, 115 S. 1400 E, Salt Lake City, UT 84112)

Particle imaging velocimetry (PIV) techniques were employed to investigate the flow patterns inside a 2 kHz annular thermoacoustic device. This thermoacoustic device uses air at one atmosphere as its working fluid and smoke for seeding particles. For these measurements, the sound level output of the device was kept to 140 dB to help maintain the suspension of seeding particles. Several types of flow patterns were observed, including Gedeon streaming, start-up noise characterized by instabilities such as vortices, and a traveling wave throughout accompanied by radial standing waves localized near corners. Such measurements are challenging in small high frequency devices. In order to satisfy the Shannon-Nyquist theorem, a 2 kHz engine would need to be sampled at a minimum of 4 kHz; however, a new technique was developed that allows for the measurements to be made using a camera with a maximum frame rate per sec of 300. This new method makes it possible to perform measurements on a flow without the need for more expensive and delicate high speed equipment, thereby, making observation of high speed flows more readily available. [Research supported by the U.S. Army Space and Missile Defense Command/U.S. Army Forces Strategic Command.]

11:45

2aPAb6. Coupling of midaudio frequency thermoacoustic prime movers. Brenna G. McDonald and Orest G. Symko (Dept. of Phys., Univ. of Utah, 115 S. 1400 E., Salt Lake City, UT 84112)

Multiple unit midaudio frequency thermoacoustic engine arrays have been studied in terms of synchronized self-sustaining oscillators. Thermoacoustic engines can be acoustically coupled through shared gas when attached to a common cavity. The coupling between devices was studied by varying the cavity volume, and by varying the separation distance between devices in an array. Frequency and phase measurements for two-engine arrays have suggested that the engines entrain at a common frequency with approximately zero phase difference implying the onset of synchronization. Amplitude measurements of synchronized arrays suggest that the acoustic amplitude is equal to the sum of the amplitudes of each uncoupled engine, and agree with predictions. Synchronization of four or more units in an array implies that the coupling between engines is global and depends on the volume of the cavity and the density of the gas within the cavity. The detuning between two engines was increased until synchronization no longer occurred. The threshold for this was found to be for a frequency difference of between 80 and 98 Hz for the given coupling strength when operating near 2.5 kHz. [Research supported by the U.S. Army Space and Missile Defense Command/U.S. Army Forces Strategic Command.]

TUESDAY MORNING, 19 MAY 2009 GALLERIA SOUTH, 8:30 TO 11:45 A.M.

Session 2aPP

Psychological and Physiological Acoustics: Theory Construction in the Domain of Auditory Perception

Frederick J. Gallun, Cochair
National Ctr. for Rehabilitative Audiology Research, Portland VA Medical Center, Portland, OR 97239-2964

Nathaniel I. Durlach, Cochair
Research Laboratory of Electronics, Massachusetts Inst. of Technology, Cambridge, MA 02139

Chair's Introduction—8:30

Invited Papers

8:35

2aPP1. Shampooing, skydiving, and the cochlear amplifier. Christopher A. Shera (Eaton-Peabody Lab., 243 Charles St., Boston, MA 02114, shera@epl.meei.harvard.edu)

What is the evidence that the mammalian cochlea amplifies traveling waves? Although mechanical nonlinearity and physiological lability are both widely cited as such, neither constitutes evidence for actual power amplification. In fact, the only real evidence comes from cochlear modeling. In this presentation I describe two modeling strategies, "shampooing" and "skydiving;" discuss the relative strengths and weaknesses of the two approaches (more conventionally dubbed the forward and inverse method, respectively); and illustrate their application to cochlear mechanics. Skydiving with a parachute provides compelling evidence that the mammalian cochlea operates as a biological, hydromechanical analog of a laser amplifier. [Work supported by the NIH/NIDCD.]

8:55

2aPP2. Development of a revised loudness model. Brian C. J. Moore (Dept. of Experimental Psychol., Univ. of Cambridge, Downing St., Cambridge CB3 9LG, UK, bcjm@cam.ac.uk)

The loudness model developed in my laboratory was based on models developed by pioneers such as Fletcher and Zwicker. Their models involved calculation of an excitation pattern for the sound of interest, from the long-term average spectrum of the sound, transformation of the excitation pattern to a loudness pattern (specific loudness plotted as a function of frequency on a perceptually based

frequency scale), and calculation of loudness from the area under the loudness pattern. The innovations used in the Cambridge model were: (1) Incorporation of a linear filter as a first stage, to account for the transmission of sound through the outer and middle ear; (2) Calculation of excitation patterns from extensive measurements of auditory filter shapes, using the notched-noise method; (3) Transformation of the frequency scale to the ERB_N-number scale, which is similar conceptually to the Bark scale, but is different numerically, especially at low frequencies; (4) Transformation of excitation to specific loudness using a function such that loudness has a finite value at the absolute threshold, as opposed to zero loudness. More recently, the model has been modified to include the concept of binaural inhibition, to accommodate experimental evidence indicating that loudness does not simply sum across ears.

9:15

2aPP3. Auditory perception through the cortical analysis of spectrotemporal modulations. Shihab Shamma (Dept. of Elect. and Comput. Eng., Inst. for Syst. Res., Univ. of Maryland, College Park, MD 20742)

To understand the functional organization of the primary auditory cortex (A1), a broad array of stimuli have been used over the years ranging from simple tones to species-specific vocalizations. In our experience, broadband stimuli with spectrotemporally modulated spectra have been most effective in evoking vigorous cortical responses and facilitating measurements of spectrotemporal receptive fields (STRFs). Extensive measurements with such stimuli led us to conclude that A1 cells analyze the spectrotemporal modulation content of their input stimulus through a rich assortment of STRFs spanning the dynamically slow to fast, the spectrally narrow to broad, and tuned at all best frequencies. These experimental findings have subsequently been formulated as a "multiresolution cortical model" whose output is postulated to represent the perceptually significant elements of sound. I shall briefly describe how complex cortical responses have been abstracted and interpreted in the context of this model, and how the model has been effective in a wide variety of applications including the assessment of speech intelligibility, representation of phonemes, explaining the perception of monaural phase sensitivity, detecting speech versus nonspeech sounds, quantifying musical timbre, enhancement of noisy speech, and in the analysis and organization of complex auditory scenes.

9:35

2aPP4. Binaural interaction modeling based on physiology. H. Steven Colburn (Dept. of Biomed. Eng., Hearing Res. Ctr., Boston Univ., 44 Cummington St., Boston, MA 02215, colburn@bu.edu)

Since psychophysical performance is based on information encoded by peripheral sensory receptors, this encoding and the resulting internal representation of sensory information impose limitations on psychophysical abilities. When enough knowledge and data are available to characterize the activity of the neurons carrying the relevant information, it is possible to derive implications for performance and constraints on mechanisms of performance. This approach has been applied for many years to multiple sensory modalities including vision and audition [e.g., Siebert, Kybernetik (1965)]. In this presentation, the impact of these ideas on understanding binaural processing will be discussed. As argued in many talks in this session, explicit models with the potential for quantitative predictions are an effective means to represent current understanding to explore hypotheses. [Work supported by NIDCD, grant No. DC-00100.]

9:55—10:10 Break

10:10

2aPP5. Autocorrelation models of pitch perception. William A. Yost (Speech and Hearing Sci. , ASU, P.O. Box 870102, Tempe, AZ 85287-0102, william.yost@asu.edu)

In 1951, Licklider first suggested autocorrelation as a method to model pitch perception in his Duplex and Triplex Theories. It was not until 1991 that a full-fledged autocorrelation model was developed by Meddis and Hewitt. Over the past decades several investigators have refined autocorrelation or autocorrelation-like models to account for pitch perception and processing. While there have been a few arguments against autocorrelation for explaining pitch phenomena, autocorrelationlike models have been very successful in accounting for a wide range of data including those associated with the pitch of iterated ripple noise. This presentation will review the development of autocorrelation models of pitch perception, especially in accounting for the pitch and pitch strength of iterated ripple noise. The presentation will also cover how refining well-established models or theories may offer a good opportunity for developing skills to build additional models and theories. [Work supported by a grant from NIDCD.]

10:30

2aPP6. Theories of hearing. Neal F. Viemeister (Dept. of Psych., Univ. of Minnesota, Minneapolis, MN 55455, nfv@umn.edu)

One of the more compelling and important aspects of auditory science has been the ability to relate perception to physiology and physics. This challenging endeavor has attracted fine minds from various disciplines and has provided theoretical frameworks for rigorously studying and describing at various levels many basic aspects of hearing. An example is the application of "signal detection theory" and its collection of tools for empirical and theoretical investigation of certain aspects of audition. But what of the future? Do we have comparable tools to construct quantitative theories that enable prediction of complex aspects of audition? Do we need new tools? Definitive answers may or may not be provided during this session.

10:50

2aPP7. Relevance of psychoacoustical theory. William Hartmann (Dept. of Phys. and Astron., Michigan State Univ., 1226 BPS Bldg., East Lansing, MI 48824)

Historically, the principles of psychoacoustics have been described in terms of mathematical theories. Are such mathematical theories now obsolete? Have they been replaced by computer algorithms? Do contemporary answers now come from simulations in computational toolboxes? The case for psychoacoustical theory can be made anew: Mathematical psychoacoustical theory has value because

of the transparency of its assumptions and the universality of its language. It has an unparalleled ability to inspire extensions and revisions in order to evolve. It illuminates new conceptual directions through analogies with other sciences subserved by similar mathematical structures. Theory consolidates large amounts of otherwise unrelated data and serves as a guide to future experiments. Mathematical theory is the practical realization of images about the way the auditory world works, images that promote understanding because of their intuitive appeal. Theory is at its best when the ordinary operations of mathematics transform an initial model into logically inescapable new insights unlikely to have arisen by other avenues. [Work supported by the NIDCD, Grant DC-00181.]

11:10—11:45 Panel Discussion

TUESDAY MORNING, 19 MAY 2009 FORUM SUITE, 8:40 TO 11:55 A.M.

Session 2aSA

Structural Acoustics and Vibration: Computational Structural Acoustics

Kuangcheng Wu, Chair
Signatures & Hydrodynamics, Northrop Grumman Newport News, Newport News, VA 23607-2770

Invited Papers

8:40

2aSA1. A three-dimensional Wiener–Hopf technique for general bodies of revolution, Part I: Symmetrization and factorization of the integral equation's kernel. Rudolph Martinez and Carina Ting (CAA-Alion Corp., 84 Sherman St., Cambridge, MA 02140, rmartinez@alionscience.com)

This is the first of a two-part paper on the development of analytic solutions of acoustic scattering and/or radiation from arbitrary bodies of revolution under heavy fluid loading. The approach followed is the construction of a 3-D Wiener–Hopf technique with Fourier transforms that operate on the finite object's arclength variable (the object's practical finiteness comes about, in a Wiener–Hopf sense, by formally bringing to zero the radius of its semi-infinite generator curve for points beyond a prescribed station). Unlike in the classical case of a planar semi-infinite geometry, the kernel of the integral equation is nontranslational and therefore with independent wavenumber spectra for its receiver and source arclengths. The solution procedure applies a symmetrizing spatial operator that reconciles the regions of $(+)$ and $(-)$ analyticity of the kernel's two-wavenumber transform with those of the virtual sources. The spatially symmetrized integral equation is of the Fredholm second kind and thus with a strong unit "diagonal"–a feature that makes possible the Wiener–Hopf factorization of its transcendental doubly transformed kernel via secondary spectral manipulations. The second of this two-part paper describes how the theory addresses departures from axisymmetry through substructuring techniques and demonstrates the analysis for canonical problems of fluid-structure interaction.

9:00

2aSA2. Midfrequency acoustic radiation analysis. Stephen D. O'Regan (Naval Surface Warfare Ctr., Carderock Div. 9500 MacArthur Blvd., West Bethesda, MD 20816, Stephen.ORegan@navy.mil)

Boundary, finite, and infinite element methods have proven useful for solving the 3-D acoustic wave equation at low frequencies. For high frequency, localized formulations such as plane-wave and ray-tracing methods have been applied successfully. These methods are impractical, inaccurate, or both for treating midfrequency problems. Low-frequency methods make very large demands on computer processing while high-frequency methods account only for local effects. Full-spectrum methods have been developed which apply interpolation between the low and high ends of the frequency range to estimate the response in the midfrequency. While these methods have provided some insight to trends as one moves toward the midfrequency range, it is not well established that they can reliably solve problems well away from their established limits. For surfaces with features having dimensions close to midfrequency wavelengths, the physics is distinct from those that govern low and high frequencies, and is not merely a linear combination of the two. This presentation reports on research to develop a midfrequency analysis tool that directly addresses the midfrequency physics. The author will identify the challenges in a physics-based method, discuss an approach to address them, and report on simulations exercising the approach.

9:20

2aSA3. The discontinuous enrichment method for multiscale and higher-frequency wave propagation problems. Charbel Farhat and Radek Tezaur (Dept. of Aeronaut. and Astronaut., 496 Lomita Mall, Stanford Univ., Stanford, CA 94305-4035)

Wave propagation problems pertain to many technologies including sonar, radar, geophysical exploration, medical imaging, nondestructive testing, and structural design. In the medium frequency regime, the analysis of these problems by the standard finite element method is either computationally unfeasible or simply unreliable, because of the well-known pollution effect. In practice, the higher-order (or *p*-type) finite element method alleviates this effect, but only to some extent. Alternative approximation methods based on plane waves have recently emerged to address this issue. The discontinuous enrichment method (DEM) is such an alternative. It distinguishes

itself from similar approaches by its ability to evaluate the important system matrices analytically, thereby bypassing the typical accuracy and cost issues associated with high-order quadrature rules. DEM also provides a unique multiscale approach to computation by employing fine scales that contain solutions of the underlying homogeneous partial differential equation in a discontinuous framework. The theoretical and computational underpinnings of this method will be overviewed. Then, recent applications to underwater acoustic and elastoacoustic scattering as well as automotive structural vibrations in the midfrequency regime will be discussed. Accuracy of 1 to 2 orders of magnitude and/or solution time improvement over the higher-order Galerkin method will be demonstrated in three dimensions.

9:40

2aSA4. Eigenanalysis in structural acoustics. Jeffrey L. Cipolla (Weidlinger Assoc. Inc., 375 Hudson St., New York, NY 10014)

Determination of characteristic frequencies and solutions (mode shapes) for vibrating systems is a staple of linear dynamics in engineering, so it should be no surprise that analysts strive for the same in coupled structural-acoustic systems. Computation of these eigensolutions confers not only physical insight into the behavior of the system, but also a means to reduce the mathematical dimension of frequency-response and other engineering computations. As in other mechanical vibration problems, the eigensolutions of coupled systems form a nearly optimal space for the approximation of the forced-response solutions. However, the coupled structural-acoustic eigenproblem is not trivial, and can be prohibitively expensive if not handled properly; this is particularly true for exterior problems. Notably, identification of a physically pertinent self-adjoint operator for the system of interest is a critical challenge. Here we will discuss the coupled structural-acoustic eigenproblem, various approximations common in practice, as well as alternative approaches to model reduction.

10:00

2aSA5. Recent innovations and current performance of conventional boundary element computations. John B. Fahnline (Water Tunnel Bldg., The Pennsylvania State Univ., University Park, PA 16802)

As long as boundary element methods have existed, research has been ongoing to make the calculations faster and applicable for larger problems. The rapid rise in computing power, both in terms of processor speed and storage capacity, has increased the problem size that can be addressed and has led to many novel strategies for reducing the computational burden. For example, many of the matrices can be stored at coarse frequency spacing over the range of interest and interpolated at much finer resolution, reducing the computation times enormously. Other strategies for speeding the computations that will be discussed include condensed acoustic meshes, optimized math kernel libraries, and multipole expansions. Results will be presented to indicate typical computation times and problem size limitations.

10:20—10:40 Break

Contributed Papers

10:40

2aSA6. Sonic boom propagation into urban landscapes: Preliminary study. Kimberly Lefkowitz (CSTI Acoust., 15835 Park Ten Pl. Ste. 105, Houston, TX 77084, kim@cstiacoustics.com) and Victor W. Sparrow (Penn State Univ., University Park, PA 16802)

In order to characterize the sonic boom response in the interiors of buildings in urban settings, first the loading forces impinging on such buildings must be understood. To do this in an urban setting, many factors must be taken into account including the shape and surface structure of the building of interest, the kind of boom being propagated, and the scale of the propagation space. Nearby buildings and their features also require consideration. In addition, the importance of specular reflection versus diffuse reflections should be addressed. In this study, these factors are discussed and evaluated relative to the impact on the overall sound field. Preliminary results on a simple case are presented. [Work supported by NASA.]

10:55

2aSA7. Wideband fast multipole accelerated boundary element methods for the three-dimensional Helmholtz equation. Nail A. Gumerov and Ramani Duraiswami (Perceptual Interfaces and Reality Lab., UMIACS & Comput. Sci. Univ. of Maryland, College Park, MD 20742)

The development of a fast multipole method (FMM) accelerated iterative solution of the boundary element method (BEM) for the Helmholtz equations in three dimensions is described. The FMM for the Helmholtz equation is significantly different for problems with low and high kD, where k is the wave number and D the domain size, and for large problems, the method must be switched between levels of the hierarchy. The BEM requires several approximate computations: numerical quadrature and approximations of the boundary shapes using elements. These errors must be balanced against approximations introduced by the FMM and the convergence criterion for an iterative solution. These different errors must all be chosen in a way that, on the one hand, excess work is not done and, on the other, that the error achieved by the overall computation is acceptable. Details of translation operators for low and high kD choice of representations and BEM quadrature schemes, all consistent with these approximations, are described. A novel preconditioner using a low accuracy FMM accelerated solver as a right preconditioner is also described. Results of the developed solvers for large boundary value problems with $0.0001 \leqslant kD \leqslant 500$ are presented and shown to perform close to theoretical expectations.

11:10

2aSA8. Energy flux streamlines vs the alternatives for the visualization of energy coupling at and inside the surface of an ensonified fluid loaded elastic cylindrical shell with vacuum interior: Preliminary results. Cleon E. Dean (Phys. Dept., P.O.B. 8031, Georgia Southern Univ., Statesboro, GA 30460-8031, cdean @GeorgiaSouthern.edu) and James P. Braselton (Georgia Southern Univ., Statesboro, GA 30460-8093)

Energy flux streamlines are compared with alternative presentations of the energy flux vector field to visualize energy coupling at the surface and inside an ensonified fluid loaded elastic cylindrical shell with vacuum interior for both forward and retrograde guided propagating waves. The present work uses a method adapted from a simpler technique due to Kaduchak and Marston [Gregory Kaduchak and Philip L. Marston, "Traveling-wave decomposition of surface displacements associated with scattering by a cylindrical shell: Numerical evaluation displaying guided forward and backward wave properties," J. Acoust. Soc. Am. **98**, 3501–3507 (1995)] to isolate unidirectional energy flows.

11:25

2aSA9. Locating arbitrary noise sources in three-dimensional space in real time. Na Zhu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202 dv6930@wayne.edu) and Sean Wu (Wayne State Univ., Detroit, MI 48202)

A portable, easy to use and effective device using four microphones only is developed for locating arbitrary noise sources in 3-D space in real time. The microphones can be placed anywhere in space so long as they are not on the same plane. This provides application flexibilities. Using cross-correlations, the differences in arrival times of these four microphone channels can be specified. This information together with a spherical spreading assumption of acoustic wave traveling in a free field enables one to determine the (x, y, z) coordinates of target sources. Error analyses are conducted and the impacts of signal noise ratios, microphone spacing, source distances, and frequencies on the accuracy of localization are examined. In particular, an empirical formulation for estimating the errors in locating a target source is developed. Using this empirical formulation, one can locate a target source for any given signal noise ratio, microphone spacing, source distance, and frequency range. Numerical simulations and experimental validations are conducted to check the effectiveness of this device to locate a variety of real world sounds, including truck pass-by noise, thundering, human voices, etc. Satisfactory results are obtained in most cases, even when the sources are behind the measurement microphones.

11:40

2aSA10. Low frequency radiation from a (compact) structure with and without sound holes. Kyle Martini (84 Sherman St., Cambridge, MA 02140, kmartini@alionscience.com)

At long wavelengths, a force driven, closed elastic structure radiates as a dipole. However, this may or may not be the case in the presence of a sound hole. Radiation remains dipole-like when the interior volume behaves as an acoustic fluid coupled to the structure. This is commonly referred to as the sound hole sum rule in musical acoustics [Weinreich (1985)]. However, should the interior acoustic volume be uncoupled, or simply ignored, the radiated field is dominated by a monopole contribution. For example, this may occur with mechanically isolated, piping systems exposed to the exterior medium. In this situation, the monopole or dipole nature of the radiation depends on the effective number of sound holes, for example, whether the driven system is well connected to the exterior along both the inlet and outlet. Finite element structural-acoustic models are developed and exercised to investigate these systems and issues. In addition, a substructuring technique is utilized to account for the potential influences of interior structural complexity.

TUESDAY MORNING, 19 MAY 2009 GRAND BALLROOM I, 8:00 A.M. TO 12:00 NOON

Session 2aSC

Speech Communication: Phonetics and Phonology (Poster Session)

Ratree Wayland, Chair
Program in Linguistics, Univ. of Florida, Gainesville, FL 32611-5454

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 even-numbered papers will be at their posters from 10:00 a.m. to 12:00 noon

2aSC1. The perception of word-final nasals in Southern Min and Mandarin Chinese. Ying Chen and Susan G. Guion (Dept. of Linguist., 1290 Univ. of Oregon, Eugene, OR 97403, ychen12@uoregon.edu)

There is a historical pattern of word-final nasal merger in Chinese languages. Some languages preserve the old Chinese final nasals /m, n, ŋ/, others preserve only /n, ŋ/, while others preserve only /ŋ/. The pattern of merger and loss, /m/ > /n/ > /ŋ/, may be due to a hierarchy of confusability among the nasals. Here, relative confusability of nasals was examined in identification tasks. Effects of vowel quality and lexical tone were also considered. First, perception of final nasals in the Quanzhou dialect of Southern Min, which preserves /m, n, ŋ/, found that /m/ was most often confused, while /n/ was least often confused. High vowel /i/ conditioned more mishearing than mid vowel /ə/ and low vowel /a/; low-level tone conditioned more mishearing than high-level, midlevel, midlow-rising and midhigh-falling tones. Second, perception of final nasals in Mandarin, which preserves /n, ŋ/, found that /n/ was more often confused than /ŋ/. High vowel /i/ again conditioned more mishearing than /ə/ and /a/; low-dipping tone conditioned more mishearing than high-level, midrising, and high-falling tones. The results do not strongly support the hierarchy of confusability among the nasals but suggest that high vowels and low tones condition more misidentifications.

2aSC2. Velarization of word-initial /l/ in African American English. Janneke Van Hofwegen and Erik R. Thomas (Dept. of English, North Carolina State Univ., Box 8105, Raleigh, NC 27695-8105, ethomas@social.chass.ncsu.edu)

Word-initial /l/ was examined in African American English in North Carolina, with comparisons with European American English, archival interviews with ex-slave African Americans and Spanish. The degree of velarization was measured using the F2-F1 difference in Bark. Interactions with formant values of the following vowel and the duration of /l/ were examined. Past research has produced conflicting results. First, initial /l/ in English is regarded as relatively unvelarized, but some research has found it to be velar. Second, there is disagreement about whether velar or nonvelar /l/ shows greater duration. Third, there has been controversy over whether velarization of /l/ is categorical or scalar. Results showed that English speakers exhibited significantly more velarization than Spanish speakers. Moreover, the degree of velarization increased across generations of African Americans. Less velarized /l/ proved to show greater duration than more velarized /l/. Velarization was quite scalar: there was no discernable division between more and less velarized tokens. /l/ tended to be more velar before a back vowel than before a front vowel, but this correlation showed no interaction with the observed generational and individual differences in velarization. [Work supported by NSF.]

2aSC3. An acoustic study of /w/ in Korean. Sun-Ah Jun and Jason B. Bishop (Dept. of Linguist., UCLA, Los Angeles, CA 90095-1543)

This study examines acoustic characteristics of Korean /w/. Modern Korean is claimed to have four /w/ diphthongs: /w/ before two front vowels (/wi, we/) and two back vowels (/wa, wʌ/). The two front diphthongs are a product of a more recent sound change from front rounded monophthongs, high /y/ and mid /ø/. However, a pilot study shows that the phonetic real-

izations of /w/ before front vowels differ from those before back vowels, and also differ whether /w/ comes after a consonant (CwV) or not (wV). It was found that /w/ before front vowels is similar to a labial-palatal approximant, especially after a consonant, showing a trace of front rounded monophthongs. To confirm this finding, a large set of data was collected from six Korean speakers and the acoustic realizations of Korean /w/ were compared with those of French labial-palatal and labial-velar approximants produced by six French speakers. The implication of the results in the phonological system will be discussed.

2aSC4. Sound change and perceptual weight of acoustic correlates to Korean stops. Kyoung-Ho Kang and Susan Guion (Dept. of Linguist., Univ. of Oregon, 1290 Univ. of Oregon, Eugene, OR 97403)

The Korean stop system is experiencing a sound change. Younger Korean speakers born in the 1970s or after produce aspirated and lenis stops differently from older speakers. Results from clear speech production experiments revealed differences in clear speech enhancement patterns. Older speakers enhanced voice onset time (VOT) difference, whereas younger speakers primarily enhanced F0 difference [K.-H. Kang and S. G. Guion, J. Acoust. Soc. Am. **124**, 3909–3917 (2008)]. These results indicate that older and younger speakers have different phonetic targets for the production of stops. The primary goal of the current study was to examine any perceptual differences related to the production differences. In order to assess the extent to which acoustic correlates to the stops contribute to the identification of the Korean stops, a perception experiment using resynthesized stimuli was conducted. Younger ($n=$ 10, mean of 20.2 years old) and older listeners ($n=$ 10, mean of 48.5 years old) participated in a stop identification task. The results suggested that F0 is a more weighted correlate than VOT for younger listeners for the distinction of the aspirated and lenis stops.

2aSC5. Phonation categories in Cajonos Zapotec. Laura Tejada (Dept. of Linguist., Univ. of So. California, GFS 301, Los Angeles, CA 90089, ltejada @usc.edu)

Phonation categories such as breathy, modal, and creaky have been described using the acoustic measurements (H1–H2) and (H1–A1). The measure (H1–H2) is the difference in dB between the first and second harmonics and is used to approximate the proportion of a glottal cycle in which the glottis is open (the open quotient). (H1–A1) is the difference in dB between the first harmonic and the most prominent harmonic in the F1 region. (H1–A1) is used as a measure of F1 bandwidth that is related to degree of glottal opening. These measures were applied to an undocumented variety of Villa Alta Zapotec spoken in San Miguel Cajonos. Acoustic data from four speakers were analyzed. Measures were taken at four evenly spaced points in time for /i/, /e/, /a/, and /o/. Results show that (H1–H2) and (H1–A1) discriminated correctly between the categories breathy, modal, and creaky with varying percentages of accuracy at different time points. (H1–H2) had a (58%) accuracy rate at time point three, compared with (63%) for (H1–A1). Combining the two measures showed that the accuracy rates rose from point one (50%) through points two (61%) and three (65%), and fell slightly at point four (58%).

2aSC6. The phonetics of contrastive phonation in Gujarati. Sameer ud Dowla Khan (Pitzer College, 1050 N. Mills Ave., Claremont, CA 91711, and Dept. of Linguist., 3125 Campbell Hall, UCLA, Los Angeles, CA 90095) and Victoria Thatte (UCLA, Los Angeles, CA 90095)

Gujarati is known for distinguishing breathy and modal phonation in both consonants and vowels [(Cardona and Suthar, 2003)], e.g., /ba:r/ "twelve" versus /Ba:r/ "burden" versus /bA:r/ "outside," where uppercase represents breathiness. The current study investigates acoustic and glottographic properties of this three-way contrast. [Fischer-Jørgensen (1967)] and [Bickley (1982)] showed that the H1-H2 measure can reliably distinguish breathy and modal vowels, and [Esposito (2006)] further found that Gujarati listeners exclusively attend to H1-H2 to categorize audio samples from other languages, even when H1-H2 did not reliably distinguish contrastive voice quality in those languages. However, do other acoustic properties also distinguish breathiness in Gujarati? In addition what physiological differences underlie the acoustic differences? The current study tests the reliability of several phonetic measures (e.g., H1-H2, H1-A1, CPP, CQ, etc.) to distinguish 33 (near-) minimal sets contrasting breathy and modal segments in audio and electroglottographic data produced by both male and female Gujarati speakers. [Work supported by NSF].

2aSC7. The effect of boundary recoverability on geminate length in English. Grace E. Oh and Melissa A. Redford (Dept. of Linguist., 1290 Univ. of Oregon, 1451 Onyx St., Eugene, OR 97403)

[Payne (2005)] argues that fake geminates are longer than true geminates, and [Ridouane (2007)] shows that their preceding vowel duration is also significantly longer. These effects are likely due to the boundary that distinguishes fake from true geminates. If this is correct, then we might expect that boundary strength will also affect geminate realization. We investigated this possibility by comparing the absolute and relative durations of assimilated and concatenated word-internal and cross-word boundary fake geminates in English (e.g., immoral versus unnamed versus fun name). Eight speakers produced five repetitions of 32 stimuli (four types: three geminate, one singleton) in two speech styles. The results showed that although cross-boundary geminates were longer than word-internal concatenated and assimilated geminates in absolute terms, both types of concatenated geminates were shorter than assimilated geminates in relative terms. A follow up experiment, comparing geminates in compound words to those emerging across words showed no differences in relative duration between the two. These results suggest that boundary strength may be less important than boundary recoverability in production. Overall, we argue that boundary recoverability impacts the phonetic implementation of geminates and likely does so whether these are true or fake.

2aSC8. Articulatory comparison of Tamil liquids and stops using real-time magnetic resonance imaging. Michael Proctor (Yale Univ., New Haven, CT 06520), Louis Goldstein, Dani Byrd, Erik Bresch, and Shri Narayanan (Univ. of Southern California, Los Angeles, CA 90089)

It has been hypothesized that liquid consonants differ from stops in that they exhibit more global control over tongue shapes. This has been demonstrated by ongoing work showing greater vowel context effects in the production of stops than liquids in Spanish and Russian. Languages with rich inventories of coronal consonants, like Tamil, provide an interesting challenge for this view, as the presence of retroflex-dental contrasts among the stops may require substantial global tongue control, and therefore inhibit context effects. To test this, dynamic magnetic resonance imaging (MRI) was used to investigate the midsagittal production of Tamil liquids in intervocalic environments. MRI sequences were acquired at a frame rate of 22.4 Hz. The five liquids of Brahmin Tamil were elicited from three speakers in three different vowel contexts. Nasal stop consonants were also elicited at the same places of articulation to compare the influence of vocalic coarticulation on liquids and stops. Greater coarticulatory influences were observed during the production of nasals than during the production of liquids in the same vowel contexts, suggesting that the production of liquids involves a more active control of dorsal articulation than that used in the production of stops. [Work supported by NIH.]

2aSC9. A study of articulatory kinematics of American English diphthongs. Sungbok Lee and Dani Byrd (Dept. of Linguist., Univ. of Southern California, 3601 Watt Way, GFS-301, Los Angeles, CA 90089)

Diphthongs are acoustically characterized by formant movements between initial and final vocalic segments [cf, Holbook and Fairbanks, J. Speech Hear. Res., (1962); Thomas Gay, J. Acoust. Soc. Am., (1968)]. It has been observed that the initial (onset) and final (offset) portions may not correspond to typical monophthongal vowels and that there exists more acoustic variability in the offset portions than in the onset portions. Spectral transition rates have been found to be different from diphthong to diphthong. In the current study, the articulatory properties of American English diphthongs are investigated in tandem with their acoustic properties. We collected both acoustic and articulatory data of five diphthongs and seven monophthongs in [b-t] context embedded in a carrier sentence using electromagnetic articulography. The data were recorded from two male and one female subjects, and each diphthong was repeated five times. The main analysis focus is on the kinematic characteristics of the three different tongue sensors

(tongue tip, tongue dorsum, and tongue rear) in each diphthong production and the comparisons to monophthongs that are close to either the onset portion or the offset portion of the given diphthong. [Work supported by NIH.]

2aSC10. Revisiting the role of the sublingual cavity in the /s/-/ʃ/ distinction. Christine H. Shadle, Michael I. Proctor, Khalil Iskarous (Haskins Labs., 300 George St., New Haven, CT 06511, shadle@haskins.yale.edu), and Maria A. Berezina (MIT-SHBT, Cambridge, MA 02139)

The sublingual cavity (SC) has been shown to play an important role in shaping the spectra of consonants, potentially distinguishing /s/ from /ʃ/. This work reports on acoustic-articulatory experiments using MRI and mechanical modeling designed to investigate the details of how the SC acts as an acoustic filter and whether it also affects the aeroacoustic source in fricative production. Five American English subjects were imaged while producing fricatives in various vowel contexts. Two subjects that show opposite patterns were studied in detail. In M1, the /s/-/ʃ/ distinction is mainly carried by SC presence and lip protrusion in /ʃ/ versus absence in /s/, resulting in marked spectral differences. M2 exhibits significant variability in SC shape and size, and tongue and lip position, in both fricatives; the acoustic spectra likewise vary. Mechanical model experiments show that the noise generation changes as the sublingual cavity is filled in. The main resonance frequency does not necessarily change. It seems thus that the sublingual cavity is just one of many articulatory variables available to the speaker and may not be the principal cause of observed acoustic differences in all subjects. [Work supported by NIH NIDCD DC 006705 to Haskins, NIH NIDCD T32 DC00038 to MIT-SHBT.]

2aSC11. End-state comfort governs kinematic variation in English flap/tap sequences. Donald Derrick and Bryan Gick (Dept. of Linguist., Univ. of British Columbia, Totem Field Studios, 2613 W. Mall, Vancouver, BC, V6T 1Z4, Canada)

Derrick & Gick [Can. Acoust. **36**(3), 162 (2008)] found that North American English flap/tap singletons are produced using up to four phonetically conditioned kinematic variants (up-flaps, down-flaps, postalveolar taps, and alveolar taps), dependent on the position of the tongue tip in neighboring rhotic and nonrhotic vowels. Our pilot work has shown that sequences of multiple flap/tap variants are highly mechanically interdependent (e.g., subjects never produce two up-flaps or two down-flaps in adjacent syllables). The present study tests flap/tap sequences for compliance with "end-state comfort" [Rosenbaum *et al.*, Acta Psychologica **94**, 59 (1996)], as evidence for look-ahead planning. Twelve subjects' tongue positions were measured during production of English words and word pairs with one- and two-flap sequences. Results show that sequences of flaps end with a preferred tongue tip position. For example, for 11 subjects, the /d/ of "edit" is realized as an alveolar tap or down-flap in "edit the," but an up-flap in "edit a," both facilitating ending the sequence with a low tongue position. A general avoidance of a sequence-final up-flap was also observed, such that /d/ of "edify" is an alveolar-tap or down-flap, while, for eight subjects, "editor" shows an up-flap, postalevolar tap sequence. [Work supported by NSERC.]

2aSC12. An articulatory study of lexicalized and epenthetic schwa using real time magnetic resonance imaging. Adam Lammert (Dept. of Comp. Sci., Univ. Southern California, 941 W. 37th Pl. Los Angeles, CA 90089), Erik Bresch, (Univ. Southern California, Los Angeles, CA 90089), Dani Byrd, Louis Goldstein, (Univ. Southern California, Los Angeles, CA 90089), and Shrikanth Narayanan (Univ. Southern California, Los Angeles, CA 90089)

Debate exists about the nature of schwa in terms of articulatory planning and execution. It is commonly believed that schwa lacks a specific constriction target. Another view is that the constriction target for schwa is the mean of the language's targeted vowels. Some have argued that both lines of thinking are correct [I. Smorodinsky, J. Acoust. Soc. Am. **109**, 2446 (2001)]. Specifically, epenthetic schwas (e.g., needed) may indeed be targetless, while lexical schwas (e.g., panda) have a target. To investigate this, we collected midsagittal realtime l MR images [S. Narayanan et al., J. Acoust. Soc. Am. **115**, 1771 (2004)] from two native English speakers. The speakers read sentences aloud which contained, variously, lexical or epenthetic schwas in multiple contexts. Images of the various schwas were extracted from the MR data and a profile of the vocal tract was fit to the image. If both targeted and targetless schwas exist, then they should be evident by examining the vocal tract shape in these images. Lexical schwas, if they are targeted, should have relatively low variability as compared to epenthetic schwas. Conversely, the vocal tract shape during epenthetic schwas should be highly correlated with the surrounding vowel context. [Work supported by NIH.]

2aSC13. Evidence of an articulatory saturation effect in the production of /s/ in American English. Joseph S. Perkell, Melanie L. Matthies, and Mark Tiede (Speech Commun. Group, Res. Lab. of Electronis, MIT, Rm. 36-591, 50 Vassar St., Cambridge, MA 02139.)

Previous work [JSLHR, *47*, 1259–1269 (2004)] demonstrated that speakers of American English who consistently activated a contact sensor on the lower alveolar ridge with the tongue tip during /s/ but not /sh/ tended to produce greater sibilant contrast than speakers who did not consistently show this contact difference. This result is compatible with the existence of a sublingual cavity during /sh/ that is eliminated by the tongue-ridge contact for /s/, which creates an abrupt decrease in volume and resulting rapid increase in resonant frequency of the cavity anterior to the sibilant constriction. This study uses data from articulatory movements (EMMA), contact of the tongue tip with the alveolar ridge, and analysis of the acoustic signal to investigate a hypothesized articulatory saturation effect: abrupt articulatory and acoustic changes result from continuous forward movement of the tongue, which does not have to be controlled precisely in stopping at /s/ onset. In tentative support of this hypothesis, preliminary observations from productions of continuous /sh/-to-/s/ transitions show movements of tongue-blade points, which slow but do not stop immediately following articulatory contact and/or a rapid rise in sibilant spectral mean. Additional tests of the hypothesis and corresponding results will be reported. [Work supported by NIDCD.]

2aSC14. Relation of vocal tract constriction location to identification of voiced stop consonants. Brad H. Story and Kate Bunton (Dept of Speech, & Lang., Hearing Sci., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721)

Consonant identification experiments are typically based on stimuli produced with a formant synthesizer where a particular acoustic characteristic is varied along some continuum of values (e.g., onset frequency of F2). In this experiment, however, vowel-consonant-vowel (VCV) stimuli were generated with a synthesizer based on a kinematic model of the vocal tract shape [B. H. Story, JASA, 117(5), 3231–3254 (2005)] where the acoustic characteristics result from the time-varying tract configuration. Each VCV was produced by superimposing a constriction at a particular location within the vocal tract on an underlying vowel-vowel sequence. Five different vowel-vowel sequences were produced, and for each, a constriction was imposed at 20 consecutive locations extending from the lips to approximately the velum. VCVs were presented in random order, ten times each, to ten listeners, using a forced choice paradigm where only the voiced English consonants /b, d, g/ were available. Listener responses indicated that the relation of formant frequencies near the consonant, to those of the preceding or succeeding vowel can provide reasonably accurate predictions of regions within the vocal tract that correspond to specific consonants. It was also revealed that the location and length of each region is dependent on the vowel. [Research supported by NIH R01-DC04789.]

2aSC15. Phonetic properties of syllabic consonants in Slovak. Marianne Pouplier (Inst. of Phonet. , Ludwig Maximilians Univ., Munich, Schellingstr 3/ii, 80799 Munich, Germany, pouplier@phonetik.uni-muenchen.de) and Štefan Beňuš (Constantine the Philosopher Univ., 94974 Nitra, Slovakia)

Studies on the articulatory correlates of syllable structure have shown that onset and coda consonants differ systematically in the timing and amplitude of their movements, and their temporal coordination with the vocalic syllable nucleus. Little is known, however, about consonants functioning as syllable nuclei. We present EMA data on Slovak, a language that allows the liquids /l/ and /r/ to be syllabic nuclei of CCC syllables. Phonologically, syllabic consonants behave like vocalic nuclei. We investigate the kinematic properties of syllabic consonants, and potential differences between vocalic and consonantal nuclei in their patterns of coordination with onsets and codas. Our data indicate that liquids in onset clusters are shorter than singletons (e.g., /r/ in "mrak" < "rak"), yet syllabic consonants (/r/ in "mrk") pattern durationally with singletons. Generally, Slovak onset clusters show little

or no overlap, which may result in an open transition schwa. For C1C2C3 syllables, C1-C2 show more overlap compared to an onset cluster, yet open transition schwas can still be observed ("mˆrk"). C2-C3 overlap less than C1-C2 in terms of movement onset, yet C1-C2 show a bigger plateau lag. These data contribute to our knowledge about the phonetic nature of syllabic consonants and articulatory correlates of syllable structure.

2aSC16. Rate conditioned variability in Western Andalusian Spanish aspiration. Benjamin Parrell (Dept. of Linguist., Univ. of Southern California, Grace Ford Salvatori 301, Los Angeles, CA 90089, parrell@usc.edu)

Western Andalusian Spanish normally shows aspiration of /s/ in coda position, but post-, rather than pre-, closure aspiration in /s/ + /p,t,k/ sequences. This has been hypothesized to be the result of gestural reorganization, with pre-closure aspiration the result of anti-phase coordination and post-closure the result of in-phase [Torreira, 9th HLS, 2007]. This experiment explores this hypothesis by leveraging the fact that fast speech can lead to instability in gestural organization and even a spontaneous switch from anti- to in-phase coordination [Turvey, Am. Psych., 1990]. Subjects produced words with the target /s/+/p,t,k/ sequences in time with an acoustic metronome that steadily increased in rate. Results show three patterns of production: consistent preaspiration, consistent postaspiration, and a transition from pre- to post-aspiration as rate increased, corresponding to anti-phase, in-phase, and an anti- to in-phase transition, respectively. Excluding the trials with consistent postaspiration, which would not be expected to transition given their stable in-phase coordination, postaspiration increases as rate increases. If there were no change in gestural coupling, we would expect the opposite pattern: as rate increases, VOT should decrease. The result thus supports the original hypothesis that postaspiration in these sequences is the result of gestural re-phasing.

2aSC17. A production study on phonologization of /u/-fronting in alveolar context. Reiko Kataoka (Dept. of Linguist., Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650)

Recently many researchers modeled sound change as the result of phonologization [Hyman (1976)], a process whereby coarticulatory perturbation of a speech sound becomes intended properties of the sound. Although phonologization is a useful concept, diagnostic criteria have yet to be proposed. By using /u/-fronting in alveolar context in American English as a case study, this paper demonstrates how production data of various speech rates may be used to differentiate phonologized speech variations from mechanically caused variations. English speakers produced /CuC/ forms where the place of both Cs was either alveolar (text context) or other (control context), in fast, medium, and slow speech rates. [Lindblom (1963)] demonstrated that the magnitude of coarticulatory influence on the vowel reduces as the vowel's duration increases. Thus, if the fronting of the vowel in alveolar context persists regardless of its duration, then this would constitute as evidence for phonologization. The results from a study with 30 speakers show that the vowels had significantly higher mean F2 values in alveolar contexts than in bilabial or zero contexts even in slow speech rate, suggesting that speakers have distinct production goals for /u/s in different contexts, or /u/-fronting has been phonologized.

2aSC18. Acoustic cues to stop coda voicing contrasts in 1 to 2-year-olds' American English. Katherine Demuth (Dept. of Cognit. and Linguistic Sci., Brown Univ., 190 Thayer St., Providence, RI 02912), Stefanie Shattuck-Hufnagel (Massachusetts Inst. of Technol., Cambridge, MA 12139), Jae Yung Song (Brown Univ., Providence, RI 02912 and MIT, Cambridge, MA 12139), Karen Evans, Jeremy Kuhn, and Miranda Sinnott-Armstrong (Brown Univ., Providence, RI 02912)

One limitation in understanding the mechanisms underlying children's variable productions has been the reliance on perception-based segmental transcription. This raises the possibility that children may make covert, acoustic distinctions that are not perceived by the transcriber. Shattuck-Hufnagel *et al.* [S. Shattuck-Hufnagel, K. Demuth, H. Hanson & S. Stevens, submitted] investigated this issue employing Stevens' [K. N. Stevens, J. Acoust. Soc. Am. **111**, 1872–1891 (2002)] feature-based acoustic model of perception/production. They found that 2,6–3,6-year-olds exhibited systematic acoustic cues to coda voicing contrasts (e.g., dog versus duck): voice-bar was more likely to precede voiced codas, whereas vowel glottalization was more likely to precede voiceless codas. The present study extended this investigation to velar codas in three 1,6–2,6-year-olds and their mothers, showing similar results: both exhibited more use of voice-bar before voiced compared to voiceless codas. For mothers, the duration of voice-bar was also longer for voiced codas, and children showed a trend in this direction. However, only mothers showed a significantly higher use of vowel glottalization before voiceless codas. Thus, although younger children also produce some acoustic cues to voicing distinctions, others take time to become adultlike. Physiological and environmental factors are discussed to explain these results. [Work supported by R01 HD057606.]

2aSC19. Segmentation of speech based on perceptual equivalence: Case of alveolar trill. Diana Stojanovic (Dept. of Linguist., Univ. of Hawaii, 1890 East-west Rd., Moore Hall 569, Honolulu, HI 96822)

Turk et al. (2006) proposed a set of rules for segmentation in durational studies. R-sounds were not included in the recommended class due to their variable realizations. The authors suggested segmentation of tap variants was straightforward but that of approximants prohibitively difficult. In this paper, we discuss the issue of segmentation of the alveolar trill in Serbian. Three native speakers read a set of minimal pair phrases in which consonants /d/, /n/, and /r/ occurred intervocalically. Phrases were matched for the position and type of pitch-accent. The four-syllable phrases produced by each speaker were judged by the native listeners as being rhythmically equivalent. When /r/ is segmented to include only the closure, vowels matched for quality, position, and accent type are significantly longer before /r/ than before /d/ or /n/. The three consonants are all voiced thus compensatory lengthening is not likely to affect the length of the preceding vowel. Based on perceived rhythmic equivalence of the examined phrases, we formulate a segmentation rule that places the start of /r/ during the vocalic portion preceding the closure. This strategy equalizes measured durations of the preceding vowels and additionally enables us to treat /r/ equivalently in all environments.

2aSC20. Non-native listener's linguistic background and use of contextual cues. Lu-Feng Shi (Dept. of Commun. Sci. and Disord., Long Island Univ.–Brooklyn Campus, One Univ. Plaza, Brooklyn, NY 11201, lu.shi@liu.edu)

Non-native listeners have more difficulty perceiving acoustically degraded English speech than native listeners. In addition, performance in non-native listeners varies greatly depending on many linguistic factors. Recognizing the importance of these factors, the current study assessed how 30 non-native listeners utilized contextual information on the reverberant speech-perception-in-noise test [Sandridge et al., AAS (2005)] based on their linguistic profiles, obtained through the Language Experience and Profile Questionnaire [V. Marian et al., JSLHR **50**, 940–967 (2007)]. Specifically, these SPIN sentences (high versus low context) were processed in two levels of multi-talker babble (signal-to-noise ratio $=+6$ versus 0 dB) and two levels of reverberation (reverberation time $=1.2$ vs 3.6 s). Significant factors in language dominance, proficiency, learning, and use were identified using correlation and multiple regression procedures. Most strikingly, factors related to reading in English (age of acquisition, age of fluency, and preference) emerged as the strongest predictors of performance regardless of the main effect (noise, reverberation, or context). Most variance in performance could be accounted for by these reading factors and two additional factors related to speaking in English (proficiency and accent).

2aSC21. Acoustic analysis of the tones and tone sandhi in Hakka Chinese. Eric Zee and Wai-Sum Lee (Dept. of CTL, City Univ. of Hong Kong, 83 Tat Chee Aven., Kowloon, Hong Kong, eric.zee@cityu.edu.hk)

This paper analyzes F0 contours of (i) the six citations tones of Hakka Chinese which consist of four long tones that occur on open syllables and closed syllables with a nasal and two short tones on checked syllables and (ii) tone contours on 36 bisyllabic compounds, representing possible permutations of the six citation tones, and 120 commonly used trisyllabic compounds. Based on the F0 data, (i) tone values are assigned to the citation tones and (ii) tone patterns on the bisyllabic and trisyllabic compounds are defined. Tone sandhi rules are formulated, summing up the phonological changes in the citation tones when juxtaposed. Phonetic tonal coarticulations are also examined. The F0 data show that the citation tones may be represented in terms of tone letter as [53 31 33 11] (long) and [5 32] (short). Major tone sandhis include [33] becoming [35] and [53] becoming

[55] before [11], [31], or [53]; and before another [31], [31] becomes [33]. Process such as [53 53 53 53] becoming [5 5 5 53] illustrates simultaneous application of tone sandhi to only tones on the non-compound-final syllables, and [5 33 11 53 31] do not undergo tone sandhi before [5] or [33].

2aSC22. Effect of tone on consonant + vowel interaction. Augustine Agwuele (Dept. of Anthropology, Texas State Univ. San Marcos, 601 University Dr., San Marcos, TX 78666)

Examining consonant + vowel coarticulation for American English [Agwuele *et al.*, 16th ICPhS, **1**, 617–620 (2007) and Lindblom *et al.*, J. Acoust. Soc. Am. **121**, 3802–3813 (2007)], the effect of speech tempo and emphatic stress on F2onset and F2mid of the following vowel were quantified and dissociated from the underlying traditional coupling of consonant and vowel. Thus, these two prosodic factors were shown to pose additional confound to CV coarticulation. Adopting the approach of these works, this present study examines VCV sequences in Yoruba, a tonal language, in order to determine whether lexical tone produces as additional confounding effect to the interaction between consonant and tone bearing vowel that could be shown to be separate from the traditionally expected CV bonding. The study shows that the changing pattern of tone configuration in VCV sequences, e.g., from VMCVM [M=midtone] to VLCVL [L=low tone] produces differing consonant + vowel interaction. The study provides acoustic evidence for a possible independence of the "source" from the "filter," i.e., the resonances of the vocal tract [Fant, G. Acoustic Theory of Speech Production (1960)].

2aSC23. The production and perception of Mon voice registers. Arthur S. Abramson, Gordon Ramsay (Haskins Labs., 300 George St., Ste. 900, New Haven, CT 06511, arthur.abramson@uconn.edu), and Theraphan Luangthongkum (Chulalongkorn Univ., Bangkok 10330, Thailand)

As a continuation of our work on voice registers [Abramson *et al.*, Phonetica **61**, 147–171 (2004); Abramson *et al.*, Phonetica **64**, 80–104 (2007)], another language, Mon, has been examined. A voice register on a syllable is a bundle of phonetic properties among which one, usually phonation type, is paramount. Fieldwork, with support from acoustics [Thongkum, in Bamroongraks *et al.*, (eds.), International Symposium on Language and Linguistics, pp. 22–50 (1998)], has yielded primarily modal voice and higher pitch for one register and breathy voice and lower pitch for the other. Our acoustic analysis has confirmed the significance of F0 height as well as that of the intensity ratios of higher harmonics and the first harmonic. The greater attenuation of the higher spectrum for breathy voice is taken as a consequence of a larger open quotient of the glottal cycle. For validation, electroglottography of the native speakers used for our acoustic analysis has been done. A control test of the robustness of the distinction in natural speech has been run along with one with synthetic speech for the efficacy of incremental changes along three dimensions, F0, spectral slope, and turbulence mixed with glottal pulsing. Final data will be shown.

2aSC24. Tone space reduction in Mandarin Chinese. Jeffrey Berry (Dept. of Linguist., Univ. of Arizona, Douglass Bldg., Rm. 200E, Tucson, AZ 85721)

Work on clear versus conversational speech has shown that speakers expand the vowel space in clear speech, which maximizes the perceptual contrast, making it easier for listeners to parse the speech signal. This study investigates how the realization of lexical tones changes for different speech tasks, and whether it is sensitive to the number of repetitions of the same word in a conversation. Native Mandarin speakers participated in a map task in which they were instructed to explain the route on the map to a partner. The first and last tokens of each target word were extracted from the dialogues and compared to the same word read from a list. Pitch and duration were normalized and quadratic curve fitting was performed on the test items. Results of statistical analyses show that contour tones from later occurrences of the target words have a smaller slope, are more linear, and tend to have pitch values closer to the overall average pitch of the speaker. In other words, all tones are realized closer to a midlevel tone when they have been previously mentioned in the conversation. In this sense, there is a contraction of the tone space for later tokens.

2aSC25. Contextual tonal variation in level tone languages. Kristine M. Yu (Dept. of Linguist., Univ. of California, Los Angeles, 3125 Campbell Hall, Los Angeles, CA 90095, krisyu@humnet.ucla.edu)

While contextual tonal variation in languages with contour tones has been well studied [Han and Kim (1974); Gandour (1992); Xu (1994)], contextual tonal variation has been less well-studied in level tone languages [Meyers (1976)]. The acoustic variation of tones due to different tonal contexts in Hausa, Bole (with two tones H and L) and Yoruba (with three tones H, M, and L) will be investigated. The description of cross-linguistic contextual tonal variation is essential for understanding the task that infants are undertaking in acquiring tonal categories in the face of variance in tonal realization. The focus of the investigation will be how the acoustic variation compares to that in languages with contour tones and if there are general patterns across languages or language-specific differences in variation.

2aSC26. Perception of Mandarin Chinese tone 2/tone 3 and the role of creaky voice. Rui Cao and Ratree Wayland (Linguist., Univ. of Florida, 4131 Turlington Hall, PO Box 115454, Gainesville, FL 32611-5454, ruicao@ufl.edu)

Previous research has established that among the four tones of Mandarin Chinese, tone 2 and tone 3 are very similar in their pitch contour shapes and are, thus, difficult to distinguish by native and nonnative listeners. On the other hand, creaky voice has been reported to be associated with tone 3 production among native speakers. However, the role of creaky voice in the perception of tone 3 has yet to be studied systematically. This study aims to investigate (1) the extent to which creaky voice is present in tone 3 production by native speakers; (2) the perceptual boundary between tone 2 and tone 3 among native and non-native speakers; and (3) the effect of creaky phonation on this perceptual boundary. Thirty native Chinese speakers and 30 native speakers of American English who are learning Chinese will participate in this study. Tone stimuli with and without creaky voice will be presented auditorily for tonal identification. The data obtained will be analyzed to examine the effect of creaky voice in the identification of tone 3 among these two groups of speakers. In addition, the native production of tone 3 will be examined for the presence of creaky voice.

2aSC27. Perceptional experiment on Mandarin tones by Mandarin, Spanish, and Cantonese speakers. Yi-Chu Ke (Dept. of Foreign Lang. and Linguist., Natl. Chiao Tung Univ., 1001 Univ. Rd., Hsinchu, Taiwan 300, ROC), Mu-Ling Teng, Chu-ting Chen, Lai-Iok Ip, Bo-hong Lu (Natl. Chiao Tung Univ., Taiwan 300, ROC), Shih-wen Chen (Natl. Tsing Hua Univ., Hsinchu, Taiwan 30013, ROC), Hsiu-Min Yu (Chung Hua Univ., Hsinchu, Taiwan 300, ROC), and Ho-hsien Pan (Natl. Chiao Tung Univ., Taiwan 300, ROC)

This paper discussed perception of three contrastive Mandarin tones: high tone (Tone1), rising tone (Tone2), and falling tone (Tone4). Based on the previous study by Hallé (2004), in which tone (Mandarin) and non-tone (French) speakers joined, our perceptional experiment employed both identification and discrimination paradigm, and categorical result of tones was expected. We recorded two Mandarin words "tu" and "wa" in those three contrastive tones, and these six words were manipulated into same duration. We then synthesized the F0 of Tone1 to Tone2, Tone1 to Tone4, and Tone2 to Tone4 in seven steps. Subjects were separated in accordance to their mother tongues, namely Mandarin, Spanish, and Cantonese. Categorical perception is obvious in Mandarin group. Cantonese speakers revealed a similar categorical pattern to Mandarin speakers, while Spanish speakers showed no categorical pattern. On the whole, Mandarin speakers had an earlier drop-down on second or third step over each pair of tone stimuli.

2aSC28. Fundamental frequency in English and Mandarin. Patricia Keating (Phonet. Lab., Dept. of Linguist., Univ. of California at Los Angeles, Los Angeles, CA 90095-1543, keating@humnet.ucla.edu)

It has been suggested that Mandarin, as a tone language, uses a larger F0 range than English. This study addresses two questions: (1) Do English and Mandarin F0 profiles differ? (2) How does the type of voice sample which is measured affect this F0 characterization and language comparison? Native speakers of English or Mandarin produced a variety of speech samples intended to elicit a wide range of F0s. Maximum phonation frequency range was measured from pitch sweeps, and speaking F0 was measured from readings of the Rainbow passage, isolated words with various intonations, and a dramatic story. Cepstral pitchtracking was used to log the entire F0 contour for each utterance, and various measures were computed. Preliminary results suggest, first, that the maximum phonation frequency ranges are the same

for the two languages; second, that in reading Mandarin speakers use a higher mean F0, but at the same time English speakers use higher maximum F0 and wider range of F0; and third, that these differences do not depend on the speech samples being compared. [Work supported by NSF.]

2aSC29. Breathing behavior in relation to prosody and linguistic structure in spontaneous speech. Peter J. Watson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455)

Hird and Kirsner [K. Hird and K. Kirsner, Brain Lang., **80**, 536–555 (2002)] reported that F0 is reset at higher values at the beginning of grammatically complete sentences during spontaneous speech. Hird and Kirsner reported that inspirations taken at the beginning of complete sentences "appeared" to be deep, but they did not make physiological measures of breathing behavior. [Watson, *et al.* J. Acoust. Soc. Am., **113**, 2812–2819 (2003)] found that when sentences were read aloud at higher values within the vital capacity they were initiated at higher F0s; however, these measures were not made during spontaneous speech and grammatical structure was not varied. For this study, we measured breathing behavior using Respitrace while three women and one man were interviewed. Posthoc, breath groups were grammatically organized according to the method used by Hird and Kirsner. On average, breath groups that comprised complete sentences were initiated from higher lung volumes and F0 was reset at higher levels at the beginning of these breath groups. Results will be discussed in relation to the interaction of linguistic structure and breathing behavior.

2aSC30. How do ordinary listeners perceive prosodic prominence? Syntagmatic versus paradigmatic comparison. Yoonsook Mo, Jennifer Cole (Dept. of Linguist., Univ. of Illinois, Urbana–Champaign, 405 N. Mathews Ave., Urbana, IL 61801), and Mark Hasegawa-Johnson (Univ. of Illinois, Urbana-Illinois, Urbana, IL 61801)

In comprehending speech, listeners are sensitive to the acoustic variation encoding the prosodic structures that mark phrasing and prominence. The present study, based on prosody transcriptions of untrained listeners (74 monolingual American English speakers), tests which acoustic feature or feature combinations cue prosodic prominence, and specifically, whether ordinary listeners perceive prosodic prominence based on changes in acoustic measures in the local context (syntagmatic comparison) or based on the value of an acoustic measure relative to the distribution of that measure across all instances of the specific phoneme in the listener's experience (paradigmatic comparison). Subjects listened to 36 short excerpts (11–25 s) of spontaneous speech from the Buckeye Corpus, and marked prominent words on a transcript. After evaluating intertranscriber agreement rates, acoustic measures (duration, intensity, subband intensities, F0 max, and formants) were extracted from stressed vowels and normalized within two domains (syntagmatic/paradigmatic). The results show that all the acoustic measures are correlated with perceived prominence and combinations of the acoustic measures account for the variability in listeners' perception of prominence. Moreover, syntagmatically normalized acoustic measures explain more of the variability, indicating that ordinary listeners perceive prosodic prominence based on local changes in acoustic measures and especially in changes in overall intensity.

2aSC31. Computational modeling of juncture strength using articulatory synthesis of prosodic gestures. Sungbok Lee, Benjamin Parrell, and Dani Byrd (Dept. of Linguist., Univ. of Southern California, 3601 Watt Way, GFS-301, Los Angeles, CA 90089)

Prosodic boundary gestures or "pi-gestures" (Byrd and Saltzman, J. Phonetics, 2003) have been introduced to model the local slowing or lengthening of articulatory gestures in the vicinity of phrase boundaries. Computational modeling of articulatory dynamics is an important tool in assessing the predicted effects of pi-gestures of varying boundary strength on constriction gestures in varying contexts. We simulate pi-gestures within the TaDA task dynamics computational model [Nam and Kim, JASA, **116**, 172–185 (2004)] and examine how functional data analysis can provide a tool for connecting articulatory lengthening with underlying pi-gesture activation strength. Specifically, the model is applied to the articulatory synthesis of two sequences: [CV#CV] and [CVC#CV], where C is bilabial or alveolar. The pi-gesture's midpoint is coordinated synchronously with the midpoint of following consonant's constriction gesture, and pi-gesture activation strength and duration are manipulated. Results indicate that pi-gesture activation strength has a much stronger effect on slowing than its duration. The slowing effect is asymmetrical, skewed earlier than the midpoint in the pi-gesture interval. After removing linear-time slowing effect (i.e., after length normalization), the slowing effect is slightly stronger in [CV#CV] than in [CVC#CV]. The strength of pi-gesture also affects spatial articulatory characteristics depending on constriction location and sequential context. [Work supported by NIH.]

2aSC32. Comparing contrastively focused and clear speech productions of color terms. Melissa A. Redford (Dept. of Linguist., Univ. of Oregon, 1451 Onyx St., Eugene, OR 97403-1290, redford@uoregon.edu), Adriano Vilela Barbosa, and Eric Vatikiotis-Bateson (Univ. of British Columbia, Vancouver, BC V6T 1Z4, Canada)

Words that are prosodically marked as focused in English are phonetically similar to clear speech words. Just as clear speech words are longer and better articulated (hyperarticulated) compared with matched casual speech words, prosodically focused words are longer and hyperarticulated compared to matched, unemphasized words. These similarities suggest that the implementational difference between prosodic focus and clear speech is merely one of scope: the same parameters are adjusted to either highlight a single element in an utterance (focus marking) or all the elements in an utterance (clear speech). To explore this idea we investigated the acoustics and jaw kinematics of producing color terms in either a casual style, a clear speech style, or with contrastive focus. Preliminary acoustic results indicate that clear speech lengthening is greater than focus-marking lengthening in some speakers. These same speakers also vary f0 more when contrastively focusing color terms than when producing them clearly, though this effect interacts with phrase position. Focused and clearly produced color terms are similarly hyperarticulated relative to casually produced terms. The results suggest that most speakers have different parameter settings for prosodic highlighting than for clear speech.

2aSC33. Acoustic correlates of rhythm in Western Cham. Kaori Ueki (Dept. of Linguist., Univ. of Hawaii, 1890 East-West Rd., Honolulu, HI 96822, kueki@hawaii.edu)

Chamic languages, which are part of the Austronesian language family, are generally known to bear lexical stress on final syllables [Thurgood, (2005)] but no instrumental studies of its prosodic features have been carried out. The current study seeks quantitative results to back up these impressionistic descriptions, and to examine how this language compares against the prosodic typology in June 2005. Syllable rimes of Western Cham, an understudied Chamic language, were measured for duration with respect to position of syllable within a phrase (initial, medial, final). The study shows that durations for final syllables were significantly longer, and that unstressed syllables are reduced in all positions. Preliminary results show that durations and pitch excursions of syllable rimes do not differ markedly with respect to phrasal position. Interactions of pitch and duration at lexical and postlexical levels will also be discussed.

2aSC34. Boundary- and prominence-related lengthening and their interaction. Argyro Katsika (Dept. of Linguist., Yale Univ., 370 Temple St., New Haven, CT 06520, argyro.katsika@yale.edu)

The durational effects of prosodic boundaries and prosodic prominence are well known, but their interaction is less well-understood. Recent studies in English [Turk and Shattuck-Hufnagel (2007); Byrd and Riggs (2008)] indicate that the two effects might be interdependent. Two acoustic experiments are presented, examining the role of boundary- and prominence-related lengthening and their interaction in Greek. The first experiment explores the effects of boundary-adjacent lengthening (conditions: no boundary, intermediate phrase, intonational phrase) and prominence (on the first, second, and third syllable away from the boundary). The second experiment examines these same effects, but post-boundary. Data from eight speakers were collected. The results from five speakers are presented. Results show that pre-boundary, both prominence and boundary have an effect on segment duration. Two prosodic levels (the level of word and one higher level) are distinguished. Furthermore, for two speakers, there is an interaction effect, such that three prosodic levels are distinguished when the prominent syllable is boundary-adjacent. The post-boundary condition shows less systematic results, with different speakers exhibiting different patterns.

These results suggest that prosodic effects are speaker dependent, more stable pre-boundary, and that prominence and boundaries interact, increasing prosodic boundary marking. Implications for models of prosodic structure are discussed.

2aSC35. Real-time MRI tracking of articulation during grammatical and ungrammatical pauses in speech. Vikram Ramanarayanan (Speech Anal. and Interpretation Lab., Dept. of Elec. Eng., RTH 320, 3710 McClintock St., Univ. of Southern California, Los Angeles, CA 90089, vramanar@usc.edu), Erik Bresch, Dani Byrd, Louis Goldstein, and Shrikanth S. Narayanan (Univ. of Southern California, Los Angeles, CA 90089)

Grammatical pauses in speech generally occur at a clause boundary, presumably due to parsing and planning; however, pausing can occur at grammatically inappropriate locations when planning, production, and/or lexical access processes are disrupted. Real-time MRI of spontaneous speech production (responses to queries like "tell me about your family," etc.) was used for seven subjects to examine the articulatory manifestations of grammatical and ungrammatical pauses (manually classified as such by two experimenters depending on the presence/absence of a clausal juncture). Measures quantifying the speed of articulators were developed and applied during these pauses as well as their immediate neighborhoods. Results indicate a consistently higher articulatory speed and spatial range for grammatical compared to ungrammatical pauses, and an appreciable drop in speed for grammatical pauses relative to their neighborhoods, suggesting that higher-level cognitive mechanisms are at work in planning grammatical pauses. On the other hand, ungrammatical pauses show only a slight decrease in articulator speed at the pause but are followed by a spurt in speed immediately after, suggesting that the planning mechanism has been reengaged only at that point, giving an indication of how much time it takes to "recover" from the perturbation of the linguistic structural integrity of the utterance. [Work supported by NIH.]

2aSC36. A corpus search methodology for focus realization. Jonathan Howell and Mats Rooth (Dept. of Linguist., Cornell Univ., 203 Morrill Hall, Ithaca, NY 14853-4701, jah238@cornell.edu)

We describe a methodology for investigating the semantic-grammatical conditioning and phonetic realization of contrastive intonation using a web harvest of particular word strings followed by grammatical and acoustic analysis. A commercial audio web search engine using speech recognition retrieved 179 MP3 files purportedly containing a token of the string "than I did." In this comparative clause fragment, contrastive focus commonly falls on the subject (she did more than I_F did), on "did" (I wish I had done more than I did_F), or following (I said more now than I did $before_F$). The 96 true tokens of "than I did" were classified into the categories "subject," "did," and "following" by grammatical and semantic criteria. For each token, 5 segment intervals were hand-annotated and more than 300 acoustic parameters extracted using a Praat script. SVM machine learning classifiers were trained that identify focus classes by acoustic criteria. On a 10-fold cross-validation test, the classifier achieves 90.2% accuracy in discriminating the dominant "subject" and "following" classes. In a listening task, human subjects achieved comparable accuracy of 90.3‘ given only the acoustic target "than I did." Stepwise logistic regression revealed measures of duration, *f*0, intensity, formants, and formant bandwidths among the significant factors.

2aSC37. Prosodic planning: Local and distant effects of phrase length on pause duration. Jelena Krivokapic (Dept. of Linguist., Yale Univ., 370 Temple St., Rm. 302, New Haven, CT 06511, jelena.krivokapic@yale.edu)

Prosodic boundaries, as manifested in pause duration, are known to be influenced by structural properties of the phrases immediately preceding and following the boundary, such as their syntactic structure, prosodic structure, and phonological length. Little is known, however, about how phrases at a distance from a boundary, i.e., not immediately preceding or following it, may affect the durational characteristics of pausing at that boundary. An experiment is presented that investigates the effects on pause duration of prosodic phrases of different length, both immediately at the boundary and farther away from it. The goal is to (a) examine local and distant effects of prosodic phrase length on pause duration and (b) to examine incrementality in production of prosodic structure, i.e., to examine how far ahead speakers plan an utterance, and how prosodic structure affects the planning process. Subjects read 24 English sentences varying along the following parameters: (a) length in syllables (two or four syllables) of the intonational phrase immediately following a target pause and (b) length in syllables of the second, distant, intonational phrase following the pause (10 or 16 syllables long). The synchronous speech paradigm [F. Cummins (2002); E. Zvonik, F. Cummins (2002)] was used to minimize variability in production.

2aSC38. Duration and fundamental frequency patterns in Brazilian Portuguese polysyllabic words. Pablo Arantes and Plinio Barbosa (Dept. of Linguist., Campinas St. Univ., PO Box 6045, Campinas, SP, Brazil 13083-970, parantes@fastmail.fm)

Brazilian Portuguese polysyllabic words are said to have secondary stress (SS). There is no clear consensus, though, on how SSs are assigned and what the acoustic correlates for SSs are. We analyzed a corpus of read speech in order to gain some insight on those two issues. The corpus' target words have 2 to 4 pre-stressed syllables with primary lexical stress falling either on the last or penultimate syllable. Target words appeared in initial, medial, or final position on carrier sentences. Previous works showed that duration and F0 are the main correlates of stress marking in BP with a marginal contribution of vowel quality and spectral emphasis. Statistical analysis of normalized (*z*-scored) syllable and vowel-to-vowel duration show no evidence of binary alternation. Duration increases exponentially up to the stressed syllable. A Praat script was used to determine global and local F0 maxima and minima. Lexically stressed syllables have F0 peaks associated with them. There is at most one F0 peak on the pre-stressed syllable chain. Peaks tend to be higher in 4-pre-stressed-syllable words and be absent in 2-pre-stressed-syllable words. Target word position do not have a significant effect on duration contour shape and on F0 peak placement.

2aSC39. Rhythmic structures in content word untterances. Joo-Kyeong Lee (Dept. of English Lang. and Lit., Univ. of Seoul, 90 Jeonnong-dong, Dongdaemun-gu, Seoul, 130-743, Korea) and Sunmi Kang (Korea Univ., Seoul, 136-701, Korea)

This study investigates rhythmic structures in the utterances composed of only content words. Content words are potentially accentable in utterance, and English necessitates a rhythmic alternation between strong and weak beats. What will happen if only content words are sequenced and they are all one-syllable words? A phonetic experiment was conducted to see if a rhythmic pattern is still produced in content word utterances and how it is manifested along the intonational contours. Results showed that merely 7% of the utterances were produced with every content word accented. Most utterances were rhythmically produced with every two (38%) or three (26%) words accented, and the other patterns such as a H^* and L^* alternation, two H^*'s intervened with L-, and consecutive downstepped H^*'s were interpreted as conforming the tonal rhythm [J.-K. Lee and S. Kang, Korean J. Speech Sci., **10**(2), 303–317 (2003)]. In two-syllable content word utterances, the patterns where two words were unaccented suggest that strong and weak rhythmic beats should be counted with the number of words rather than syllables at the level of utterance. It is proposed that it is an accentual foot as opposed to a lexical foot at the level of lexical word.

2aSC40. The production of English lexical stress by Thai speakers. Jirapat Jangjamras and Ratree Wayland (Program in Linguist., Univ. of Florida, Box 115454, Gainesville, FL 32611-5454, jirapat@ufl.edu)

The major goal of this work is to report acoustic correlates of English lexical stress produced by native and second language learners of American English. Fifteen native Thai speakers and fifteen native English speakers were asked to concatenate two English monosyllabic nonwards into one disyllabic word and to produce stress on either the first or the second of the resulting disyllables. Disyllabic English nonwords were constructed by pairing CVO, CVVO, CVV, CVS, and CVVS monosyllables. It is predicted that Thai speakers' production of English lexical stress will be influenced by Thai prosodic system including Thai tonal distribution rules which are constrained by syllabic structure, the presence of fixed final-stress placement, and the use of duration as the primary cue in L2 English lexical stress production. The acoustic parameters measured included vowel duration, intensity, fundamental frequency, and formant frequencies (F1, F2, F3). The ratio values of these measurements between stressed and unstressed syl-

lables will be correlated with stress identification scores obtained from native English listener judges to examine the amount of acoustic information required for accurate perception of stress location by native listeners.

2aSC41. Formant frequency development and gender differences in Mandarin-learning children before 5 years of age: A longitudinal analysis. Li-mei Chen (Dept. of Foreign Lang. and Lit., Natl. Cheng Kung Univ., 1 University Rd., Tainan City, Taiwan, leemay@mail.ncku.edu.tw)

This study is designed to investigate the longitudinal developmental process of vocalic production in Mandarin-learning children from birth. This current study focuses on the fifth year and the continuity in the process. Major findings are: (1) From three years and six months, the boy subject showed lower vowel formant values, reflecting gender differences in the development of vocal tract. In addition to three years and six months, there was another stage of decrease of frequency at four years and six months in the boy subject. The trend was especially evident in F2 values. (2) The girl subject used more high front rounded vowels /y/, which is the only late-acquired single vowel. (3) Corner vowels /i, u/ were used more often by the girl subject. (4) Diphthongs continued to show high frequency of occurrence at this stage. (5) The frequency of vowels with nasal endings showed prominent increase at this stage in both genders. However, the girl subject showed more preference in using this vowel category as was found in a previous stage. Gender differences in vowel development found in this stage can be traced in the following longitudinal observation. [This investigation was supported through funds from National Science Council in Taiwan, Republic of China (NSC 96-2411-H-006 -023).]

2aSC42. Effects of speaker gender on voice onset time in Korean stops. Eunjin Oh (Dept. of English Lang. and Lit., Ewha Womans Univ., 11-1 Daehyun-dong, Seodaemun-gu, Seoul 120-750, South Korea, ejoh@ewha.ac.kr)

Several studies have shown that the voice onset time (VOT) of English stops varies as a function of speaker sex. The overall trend indicates that females exhibit longer VOT values than males for voiceless stops. The suggested explanations have mainly attributed this difference to physiological or stylistic differences between the sexes. If this is so, we would expect to find identical sex differences in the VOTs of stops in other languages. This study tested 38 native speakers of Korean (19 males and 19 females), who produced fortis, lenis, and aspirated stops in /CVn/ words. The males tended to exhibit longer VOTs than females for most of the syllables, the greatest differences occurring in aspirated stops. The males also exhibited more distinctive distributions in VOT values than the females in Korean. These results indicate that the physiological or stylistic explanations previously suggested are not tenable since there is no automatic reason why VOT values for voiceless stops are, in general, longer for females in English, but longer for males in Korean. Sociophonetic factors may explain why languages can exhibit opposing trends in gender differences in temporal acoustic properties such as VOTs.

2aSC43. Variation in the phonetic realization of Thai lexical tones across speaking styles. Donruethai Laphasradakul and Ratree Wayland (Program in Linguist., Univ. of Florida, Gainesville, FL 32603)

Thai is a tonal language with five contrastive tones: mid-falling, low-falling, high-falling, high-rising, and low falling-rising. Previous literature has shown that the acoustic realization of Thai lexical tones, namely their fundamental frequency or F0, is affected by the phonetic contexts in which they are produced. For example, in connected, conversational speech, Zsiga [(2006)] found that F0 contours of Thai tones produced in phrase-final position are different from those produced in non-phrase-final position. Context-dependent variation in F0 contours among lexical tones in Thai was also noted in Luksaneeyanavin [(1983)]. The main aim of this research was to further investigate F0 variations across the five lexical tones in Thai. Ten native Thai speakers (5 males and 5 female) were asked to produce Thai words in two speaking styles: slow, careful reading style; and fast, conversational style. Preliminary analyses revealed that: (1) the high-falling tone exhibited the highest degree of variation; (2) a greater amount of variation was observed among female speakers than among male speakers; and finally (3) a greater amount of variation was observed in open syllables than in closed syllables.

2aSC44. Quantifying the time-varying coordination between performer and audience. Adriano Barbosa (Linguist., Univ. of British Columbia, 2613 West Mall, Vancouver, BC, Canada, V6T 1Z4, adriano.vilela@gmail.com), Kevin G. Munhall (Queen's Univ., Kingston, ON, Canada, K7L 3N6), Hani C. Yehia (CEFALA, UFMG, Belo Horizonte, MG 31270-901, Brazil), and Eric Vatikiotis-Bateson (Univ. of British Columbia, 2613 West Mall, Vancouver, BC, Canada, V6T 1Z4)

This study examines the coordination that occurs ubiquitously during behavioral interaction (e.g., linguistic, social, musical). Coordination, however, does not imply strict synchronization. Musicians in a quartet, for example, will deviate slightly from the measured beat, alternately playing slightly ahead or behind. This paper showcases a new algorithm for computing the continuous, instantaneous correlation between two signals, at ANY temporal offset, resulting in a two-dimensional mapping of correlation and temporal offset that makes it possible to visualize and assess the time-varying nature of the coordination. The algorithm is demonstrated through analysis of the extraordinary performer-audience coordination evoked by Freddie Mercury during the rock group, Queen's, Live Aid Concert at Wembley Stadium in 1985.

2aSC45. Perception of intonation as a function of sex. Yu-Wen Chen and Peter J. Watson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, Shevlin 115, Minneapolis, MN 55455)

Fitzsimons *et al.* [M. Fitzsimons, N. Sheahan, and H. Staunton, Brain Lang. **78**, 94–108 (2001)] reported differences in the production of intonation between men and women during speech production. A relationship between the production and perception of segmental variables has been recognized [J. Perkell *et al.*, Speech Commun. **22**, 227–250 (1997)]; however, little information is available about the relationship of prosodic perception and production. Five men and five women listened to sentences produced by one man and one woman. The fundamental frequency (F0) of each sentence was resynthesized with 22 levels of F0 variation ranging from no variation to F0 variation that was 3.5 times greater than the typical variation of a sentence. For each sentence, a listener was randomly presented one of the levels of F0 variation and asked to adjust variation up and down until they were satisfied that the variation of that sentence sounded the most "natural." Additionally, each subject read a set of sentences aloud, and the F0 variation of these productions was analyzed. Results will be discussed in relation to the production and perception of intonation within and between the men's and women's groups.

2aSC46. Perceived sexual orientation and speech style: A perceptual and acoustic analysis of intentionally clear and conversational speech. Benjamin Munson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr., Minneapolis, MN 55455, munso005@umn.edu), Sarah Hargus Ferguson, and Catherine L. Connealy (Univ. of Kansas, Lawrence, KS 66045)

Munson *et al.* [Journal of Phonetics (2006)], found a correlation between perceived speech clarity and perceived sexual orientation (PSO) of male talkers: those who had been rated to sound gay were rated by an independent group to produce speech clearly. A smaller, opposite influence was found for women talkers. Munson *et al.* hypothesized that gay male speech styles might be based, in part, on clear speech styles. One weakness of Munson *et al.*'s study was that none of the speech examined was intentionally clear. This study follows up on this finding. A new group of 29 listeners rated the PSO of four men and four women producing conversational and intentionally clear speech, recorded as part of a larger study on interspeaker differences in talker intelligibility. PSO was rated using a five-point scale. Speech style had a moderate, statistically significant influence on PSO. In contrast with previous research, both men and women were rated as gayer-sounding when producing clear styles than when producing conversational ones. The differences in mean ratings were small. This investigation, thus, partly sup-

ports Munson *et al.*'s findings. The relatively small influence of speech style on PSO likely reflects the complex, content-dependent nature of gay speech styles and PSO.

2aSC47. Sexuality and personality: The effect of pitch height and contour width on perception in Japanese. Margaret Camp (Dept. of East Asian Stud., Learning Services Bldg., Rm. 102, Univ. of Arizona, 1512 First St., P.O. Box 210105, Tucson, AZ, 85721-0105, mcamp@arizona.edu)

This work investigates how Japanese listeners judge a speaker's personality characteristics and sexual orientation. Base sentences, produced in Japanese by a male and a female native Japanese speaker, were resynthesized to manipulate specific acoustic characteristics believed to influence perceived sexual orientation or personality. Overall, the higher the fundamental frequency (f0) of the female voice, the less likely it was to be perceived as lesbian. Similarly, the flatter the intonation pattern of the female voice, the more likely it was to be viewed as lesbian. For male voices, the higher the f0, the more likely the voice would be perceived as gay. Pitch variability did not play a significant role in this perception. Finally, out of the four personality scales investigated, two showed a significant correlation to lesbian voices: masculinity and calmness. Male voices showed the same pattern, with gay voices correlating with femininity and emotionality. The influence of the sex of the listener was also investigated, but male and female listeners appear to make similar judgments about most, but not all, characteristics.

2aSC48. An acoustic analysis of /æ/ variation and its relationship to perceived sexual orientation in American English. Benjamin Munson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr., Minneapolis, MN 55455, Munso005@umn.edu), Kathleen Currie Hall, and E. Allyn Smith (Ohio State Univ., Columbus, OH 43210)

Munson *et al.* [J. Phonetics (2006)] found that 11 self-identified gay and 11 heterosexual men produced different variants of the vowel /æ/, with gay men producing lower, more retracted variants and heterosexual men producing higher, more tense variants. Listeners' performance in a perception task in which they rated these talkers' sexual orientation was correlated with these measures: talkers with higher, more-tense /æ/ were rated as sounding more heterosexual than talkers with lower, more retracted /æ/. However, Smith *et al.* [New Ways of Analyzing Variation (2008)] found the opposite pattern in an experiment in which listeners rated the sexual orientation of productions by 10 trained talkers of sentences containing either tense or retracted /æ/ variants. A different group of listeners showed the same pattern when rating the /æ/ words excised from these sentences, though these listeners replicated Munson *et al.*'s original finding when presented with the /æ/ words from the original 22 talkers. An attempt to reconcile these findings is made through detailed acoustic analysis of the /æ/ productions in the two experiments. Results underscore the importance of doing careful acoustic analyses in concert with careful phonetic transcription when conducting experiments on perception of social variables and speaker attributes.

2aSC49. Parametrization of *f*0 characteristics of register ("Shiki") in mid-western Japanese dialects. Kenji Yoshida (Dept. of Linguist., Indiana Univ., 1021 E. 3rd St, Bloomington, 47405, keyoshid@umail.iu.edu)

Although the ToBI framework has achieved great success in modeling intonation in various languages, it may not capture the nature of certain aspects of f0 movements. This study models dynamic aspects of *f*0 movement in "pitch register" contrasts observed in mid-western Japanese dialects, called "Shiki" by Japanese dialectologists. An exponential function [f0 = intercept + $b \times$ time $\times \exp(-c \times \text{time})$] was fitted to the observed f0 movement of speakers of two Japanese dialects with Shiki contrasts (Kan'onji and Kishiwada). The results show that the two contrastive Shiki types are clearly discriminated in the 3-D parameter space [intercept = baseline, b = magnitude of f0-excursion, c = inverse of peak timing]. Further analysis revealed that the two contrastive Shiki types are not reliably differentiated by the baseline. The more reliable predictors of Shiki types are the other two parameters that determine f0 movement within a word. The results are more congenial to a "dynamic" definition of Shiki system by Japanese dialectologists [e.g., Uwano, (1988)] and suggests that the analysis of Shiki as an initial f0 target [(High vs. Low, Pierrehumbert Beckman, 1988 for Osaka dialect] should be reconsidered.

2aSC50. Dialect distribution of North American English /r/. Suzanne Boyce (Dept. of Commun. Sci. and Disord., Univ. of Cincinnati, Cincinnati, OH 45267, boycese@email.uc.edu), Mark Tiede (Haskins Lab., New Haven, CT 06510), Kathy Groves-Wright (Cincinnati Veterans Affairs Medical Ctr., Cincinnati, OH 45220), Carol Espy-Wilson (Univ. of Maryland, College Park, MD 20742), Christy Holland, and Kyuran Choe (Univ. of Cincinnati, Cincinnati, OH, 45267)

It is well-known that North American English /r/ may be produced by a number of different tongue shapes. This was first reported in detail by Delattre and Freeman (1968), who examined x-ray data from 46 speakers from various regions of the United States and identified six basic tongue shapes used by speakers of "rhotic" dialects to produce /r/. It is unclear, however, whether different shapes are characteristic of particular dialects or geographic regions, or a matter of individual variation based on anatomy, developmental history, etc. In this study, we review findings from a number of published x-ray and MRI studies of tongue shapes for /r/ using speakers from various dialect regions. In addition, we characterize the variability of tongue shapes used for /r/ in an unpublished x-ray database of elderly men with normal speech from the area of Cincinnati, OH. Results show that distinct tongue shape patterns in /r/ production occur both within and across dialect regions, suggesting the primacy of subject-specific factors over dialectal effects. [Work supported by NIH.]

2aSC51. Pitch accommodation in synchronous speech. Miran Kim (Dept. of Linguist., SUNY at Stony Brook, S201 SBS Bldg., Stony Brook, NY 11720, mrkim@ic.sunysb.edu) and Hosung Nam (Haskins Labs., New Haven, CT 06511)

Synchronously read speech, in which a pair of speakers reads a text together, has been shown to reduce variability in speech tempo and unveil underlying properties of speech competence. Our study question is whether a synchronous speech condition also influences a pitch domain as well as a temporal one, which has been the main focus in studies of synchrony. We investigated how tones are modulated in the synchronous speech condition. Eight speakers of Mandarin Chinese were instructed to read a story text alone and together in pairs. Speakers also read the text together with recorded speech. Each trial was repeated five times. We measured F0 at onset, center, and offset of a resonant period for each syllable and compared read-alone to read-together speeches. Results show that (1) speakers in the synchronous speech condition tend to tune their pitch to each other such that F0 difference in pairs decreases overall; (2) they exhibit less variability in F0 across their own repetitions; (3) tonal targets in the read-together condition are achieved more stably in time. The pitch accommodation found in our study implies that speakers attend to F0 manifestation as well as tempo when F0 changes are phonologically contrastive in their language.

2aSC52. Effects of semantic context and regional dialect variation on speech intelligibility in noise. Cynthia G. Clopper (Dept. of Linguist., Ohio State Univ., 1712 Neil Ave., Columbus, OH 43210, clopper.1@osu.edu)

Speech intelligibility in noise is affected by the semantic content of the utterance; semantically predictable utterances are more intelligible than semantically unpredictable or anomalous utterances. However, dialect variation interacts with semantic predictability in production; talkers produce more dialect-specific variants in semantically predictable contexts and more standard variants in semantically unpredictable contexts. The goal of the current study was to explore the interaction between dialect variation and semantic context in perception. Participants were asked to transcribe the final word of semantically predictable, unpredictable, and anomalous sentences produced by talkers from four regional dialects of American English and mixed with speech-shaped noise. The listeners exhibited an intelligibility benefit in the predictable context relative to the other two semantic contexts, regardless of talker dialect. However, differences in intelligibility between the dialects were reduced in the unpredictable and anomalous contexts relative to the predictable context, suggesting greater online adaptation to dialect variation in the unpredictable and anomalous contexts than the predictable context. These results are consistent with the finding that dia-

lect differences may be attenuated in production in low predictability contexts, and suggest that perceptual dialect adaptation is affected by fine-grained intratalker acoustic variation.

2aSC53. Investigating perceptual measures of speech alignment: Do matching tasks make the grade? Rachel M. Miller, Kauyumari Sanchez, and Lawrence D. Rosenblum (Dept. of Psych., Univ. of California, Riverside, 900 University Ave., Riverside, CA 92521)

Speech alignment is the tendency to produce speech which sounds similar to that of a person with whom one is speaking. It is often measured perceptually using an AXB-matching task [e.g., S. D. Goldinger, Psychol. Rev. **105**, 251 (1998)]. In this procedure, naïve raters determine which of two utterances of a talker, produced prior (baseline) and after interaction with a model (e.g., shadowed), sounds more similar to that model. However, utilizing this type of perceptual measure has come under scrutiny due to its reliance on baseline utterances produced by talkers reading text. It has been suggested that AXB matching results reveal more about inherent differences between read and produced (shadowed) speech, than actual alignment. The present research addressed this concern. Experiment 1 involved a modified AXB task for which raters judged whether a subject sounded more like the model they had shadowed or another model. In Experiment 2, raters judged whether a subject sounded more similar to the model they had shadowed or to another subject who shadowed a different model. In both cases, results revealed that subjects sounded more similar to the model they shadowed. This suggests that alignment is perceivable by raters even when removing the comparison to read text. [Work supported by NIDCD Grant R01DC008957.]

2aSC54. Identification of synthetic vowels based on time-varying vocal tract area functions. Kate Bunton and Brad H. Story (Dept. of Speech, Lang., & Hearing Sci., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721)

Identification accuracy for synthetic vowel stimuli generated with static vocal tract area functions from eight speakers was recently reported (K. Bunton and B. H. Story, JASA, in press). Although vowels were identified with reasonably high accuracy, in many cases neighboring vowels were confused. In the present study, new stimuli were generated based, again, on area functions from the same eight speakers. This time, however, the temporal variation of the vocal tract shape during natural production of isolated vowels was incorporated by developing a formant-to-area mapping for each speaker based on principal components analysis. This allowed formant contours extracted from recorded vowels to be mapped onto time-varying area functions. These were coupled to a voice source model and acoustic waves were propagated with a wave-reflection vocal tract model to generate vowels. Vowels were identified by listeners using a forced choice paradigm. Results indicated that including variation in vowel tract shapes, and hence formant frequencies, along with natural vowel durations improved identification accuracy relative to the previously tested static versions. It is concluded that these eight sets of area functions generate good renditions of target vowels when changes in vocal tract shape and duration were included. [Research supported by NIH R01-DC04789.]

TUESDAY MORNING, 19 MAY 2009 PARLOR B/C, 8:25 TO 11:40 A.M.

Session 2aSP

Signal Processing in Acoustics and Underwater Acoustics: Detection and Classification of Underwater Targets I

Patrick J. Loughlin, Cochair
Dept. of Bioengineering, Univ. of Pittsburgh, Pittsburgh, PA 15261

Maya Gupta, Cochair
Dept. of Electrical Eng., Univ. of Washington, Seattle, WA 98195

Chair's Introduction—8:25

Invited Papers

8:30

2aSP1. Passive detection in deep water using the reliable acoustic path. Lisa M. Zurk (Elec. & Comp. Eng. Dept., Portland State Univ., PO Box ECE 751, Portland, OR 97201, zurkl@pdx.edu)

The detection of quiet targets in deep water environments is of increasing interest for Navy applications, and detection performance depends critically on efficiently exploiting the frequency-dependent signal and noise structure. Previous measurements in the deep ocean environment of the northeast pacific has demonstrated the significant depth and frequency dependence of the noise, which can also vary with local wind conditions [Gaul *et al.*, IEEE JOE (2007)]. The signal structure depends both on the spectrum of the target itself, but also on the propagation characteristics of the channel. In the concept of a reliable acoustic path (RAP), a receiver placed at or below the critical depth can detect shallow sources out to moderate ranges independent of near-surface water conditions or bottom interaction (hence, the identification as a "reliable" path). In this presentation, we present analysis of noise and signal structure for deep water channels from historic data, propagation modeling, and more recent experimentation. This structure is then used to compute signal excess to quantify detection performance as a function of receiver array placement and topology. Implications of the results on future Navy systems is discussed.

8:50

2aSP2. Enhancing passive automation performance using an acoustic propagation simulation. Ashley J. Llorens, Trudy L. Philip, Iman W. Schurman, and Cory R. Lorenz (Natl. Security Technol. Dept., JHU/APL, 11100 Johns Hopkins Rd., Laurel, MD 20723, ashley.llorens@jhuapl.edu)

During an at-sea encounter, signatures of interest can exhibit characteristics that differ from those observed in previously recorded data. These differences can occur due to variations in a number of factors including encounter geometry, propagation channel, and receiving sensor configuration. This paper presents a simulation technique that imposes low-frequency propagation effects on a time-domain signal using a normal mode method. High-quality, time-varying recorded signatures are used as inputs into the algorithm, which outputs band-limited time-series data for a selected geometry and environment. The output time-series are phased to simulate the time-varying pressure amplitudes that would be received by a towed array or any multielement passive sensor configuration operating in a realistic multipath environment. These capabilities enable the simulation of signatures of interest as captured under a broad range of littoral conditions by various passive sensors. These simulated data are used to augment scarce signature collections for training and assessing the performance of passive sonar automation.

9:10

2aSP3. Underwater target detection and classification using high resolution sensors and wide band sonar. Yvan Petillot, Yan Pailhas, Chris Capus, and Keith Brown (Oceans Systems Lab., Heriot-Watt Univ., UK)

In this presentation, our work on integrated target detection and identification using sonar systems is presented. Standard high-resolution imaging sonar such as side scan and synthetic aperture are used for detection. Existing methods focus on statistical techniques using Markov models and machine learning-based techniques. Results on both synthetic and real data acquired in a large variety of terrains are presented. Identification can be based on either the shape of the target or its internal structure. Techniques-based active contours for shape analysis have been developed but these techniques only work well on very high-resolution data such as SAS and can be difficult to use on cluttered terrain. We propose to use the wide-band bio-sonar we developed to perform this task. In that case, the analysis focuses on the resonances induced by the internal structure of the target and the material used. Simulation confirmed by theoretical analysis on simple targets and experimental results demonstrate the effectiveness of this approach. These techniques are integrated on real vehicles in a collaborative vehicle framework to perform detection, classification, and identification of targets. [This work was supported by Thales, The Ministry of Defence, and the European Union under contract MRTN/CT/2006/036186.]

9:30

2aSP4. What do cognitive models and human judgments suggest about the desired structure of automatic classifiers? Jason E. Summers (SAIC, 10401 Fernwood Rd., Ste. 120, Bethesda, MD 20817-1133), Charles F. Gaumond, Colin W. Jemmott, and Derek Brock (Naval Res. Lab., Washington, DC 20375-5320)

Past studies probed human listeners' efficacy at classification of impulsive sonar echoes by using paired-comparison ratings to measure perceptual dissimilarity. Interpreting these ratings requires a cognitive model comprising both a representation of the stimuli and a process operating on that representation. Initially, perceived dissimilarities were represented as distances in Euclidean space via multidimensional scaling. This assumed a continuous and spatial cognitive representation and proved difficult to relate to a linear vector space of features. Later work by the authors suggested a discrete (categorical) cognitive representation may better reflect perception of these stimuli. While similarity-based classifiers can bypasses feature extraction [S. Philips and J. Pitton, J. Acoust. Soc. Am. **123**, 3344 (2008) (A)], the form of the similarity measure reflects the assumed cognitive representation [L. Cazzanti and M. R. Gupta, Proc. IEEE Intl. Symposium Info. Theory, pp. 1836–1849 (2006)]. Here, findings based on paired-comparison ratings and verbal descriptions of stimuli are discussed in terms of the cognitive models they reflect and implications for classifier architecture: the form of similarity measures, whether features are continuous or discrete, and whether features are common to all classes. [Work supported by the Office of Naval Research.]

Contributed Papers

9:50

2aSP5. A generalized linear filtering approach for sonar receivers. Nabin S. Sharma and John R. Buck (Dept. of ECE, Univ. of Massachusetts Dartmouth, 285 Old Westport Rd., North Dartmouth, MA 02747, nsharma2@umassd.edu)

Active sonar systems operating in realistic environments face the challenge of simultaneously detecting and resolving multiple targets. Receiver signal-to-noise ratio gain and main-lobe width of the receiver response to the transmitted signal limit the detection and resolution capability of an active sonar receiver, respectively. A matched filter is commonly used for target detection to exploit its optimal detection property. Theoretically, an inverse filter achieves optimal resolution and can be used to resolve targets. In practice, the inverse filter is difficult to implement and requires unrealistically high signal-to-noise ratio to perform well. For an active sonar receiver operating in a cluttered environment with closely spaced targets, a matched filter might not resolve some targets and an inverse filter could not detect some targets. For such environments, a generalized linear filtering approach is proposed. This new approach combines the matched and inverse filter properties in a controlled manner to achieve smooth trade off between detection and resolution. In simulation, the generalized linear filter output is compared with the matched and inverse filter outputs to demonstrate the advantages of the proposed filter. In addition, receiver operating characteristics of the filters confirm the improved performance of the generalized linear filter. [Work funded by ONR Code 321US.]

10:05—10:25 Break

10:25

2aSP6. Pulse propagation and classification in time/frequency and position/wave number phase space. Patrick Loughlin (Depts. of Bioengineering and ECE, Univ. of Pittsburgh, 745 Benedum Hall, Pittsburgh, PA 15261, loughlin@pitt.edu)

In dispersive pulse propagation, different frequencies travel at different velocities, and, hence, the pulse is nonstationary in that it changes over time and distance. Joint phase space representations, such as the Wigner distribution and the spectrogram, are often employed to study nonstationary signals, and, hence, it is natural to consider their application to dispersive propagation. We consider position/wave number and time/frequency representations of a pulse propagating with dispersion and damping. We give the

Wigner distribution of the pulse at a particular time/position in terms of the Wigner distribution of the initial pulse. Approximations of this result are derived, which are accurate and remarkably simple to apply, as compared to the stationary phase approximation. The approximations provide insights as well, in that they show how each point in phase space propagates at the group velocity, and lead to new features for classification, based on local phase space moments, that are invariant to the propagation effects of dispersion and damping. Examples relevant to sound propagation and classification in shallow water are presented. [Work supported by ONR Grant No. N00014-06-1-0009.]

10:40

2aSP7. Krylov methods in time-reversal imaging by multiple-signal-classification. Paul E. Barbone (Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, barbone@bu.edu), Assad A. Oberai (Rensselaer Polytech. Inst., Troy, NY 12180), and Gonzalo R. Feijóo (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Multiple signal classification (MUSIC) has been used to form images and identify sound sources since 1986 [R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag., **AP-34**, 276–280 (1986)]. In active sonar imaging of point targets, the MUSIC method can be used to estimate the range of the time-reversal operator. In this context, the method is related to the decomposition of the time-reversal operator method. Typical implementations of these methods utilize measurements of the entire time-reversal operator, and require computations of its eigenvalues and eigenvectors. By contrast, we show that Krylov iterative methods can be used to perform MUSIC imaging with relatively few acoustic excitations. Furthermore, by using the Lanczos technique, no eigenvalues or eigenvectors need be computed. Rather, an orthonormal basis for the range space of the time-reversal operator can be constructed directly from the received data. Most of the necessary computing is performed by the array itself, performing as a kind of "analog computer." Finally, we show that useful images can be formed from one iteration to the next, while the data are being collected.

10:55

2aSP8. Automatic classification of underwater targets using fuzzy-cluster-based wavelet signatures. Hui Ou (Dept. of Elec. Engr., Univ. of Hawaii at Manoa, 2540 Dole St., Honolulu, HI 96822, ou@hawaii.edu), John S. Allen, III, and Vassilis L. Syrmos (Univ. of Hawaii at Manoa, Honolulu, HI 96822)

A novel active acoustic underwater target recognition approach is applied to simulated mine-like targets. A feature extraction and integration method is introduced based on the continuous wavelet transform (CWT) and the Gustafson-Kessel (GK) clustering algorithm. CWT projects the acoustic signals onto the time-scale space, and the time-scale signatures are represented by a number of clusters. A modified GK clustering algorithm allows for specific enhancements: (i) autoadjustment of the shape of the clusters according to the time-scale patterns and (ii) autovalidation of the optimum number of clusters. The feature integration encompasses the application of the clustering analysis N times. The resulting cluster centers are combined and overlaps are eliminated. Using this method, the target of interest is characterized by an integrated clustering distribution, or a feature template. A classification scheme is developed by matching the clustering distributions of the echoes received from an unknown target with the existing feature templates. White Gaussian noise has been added to the signals to simulate the interface of an ocean floor. The recognition rate varies from 72–98% for a SNR between 5–15 dB.

11:10

2aSP9. Distributed detection and estimation based on the turbo approach. Michael Roan and Elizabeth Hoppe (Mech. Eng., Virginia Tech., 141 Durham Hall, Blacksburg VA 24060)

We develop a centralized information fusion architecture from basic principles of information theory and Bayesian statistics. It is well known that any clustering, quantizing, or thresholding of data causes loss of information unless a sufficient statistic is computed in the processing. For the case of wideband active ranging systems, the coherent output of an optimum beamformer and a matched filter is a sufficient statistic that can be transmitted to the fusion center. For unknown target velocity, range, and bearing, the wideband space-time matched filter output can be interpreted as a multidimensional wavelet transform or a delay-scale-bearing map. In this paper, an iterative, Bayesian, joint estimation-detection approach is used for computation of sufficient statistics and multisensor information fusion. An approach borrowed from sequential Bayesian processing is used to compute prior densities for joint Bayesian estimation-detection. In this approach *a posteriori* densities become priors after a coordinate transformation that transforms the outputs of each sensor to a common reference frame for all sensors. In this paper, receiver operating characteristics and Cramer–Rao lower bounds are given for several undersea signal processing cases of interest.

11:25

2aSP10. Variational Bayesian particle filtering for underwater target localization and tracking. T. Scott Brandes, Nilanjan Dasgupta, and Lawrence Carin (Signal Innovations Group, Inc., 1009 Slater Rd., Ste. 200, Res. Triangle Park, NC 27703)

Kalman filtering methods used in target tracking applications fail to perform satisfactorily when the true underlying posterior density of a target's location is non-Gaussian. These circumstances arise often in highly noisy environments dominated by strong scattering returns from clutter. Advanced tracking approaches, such as particle filters (PFs), that do not have the linear and Gaussian requirements of Kalman filtering are needed for target tracking in those complex environments. We used a variational Bayesian (VB) particle filter in our research to accurately track underwater targets simulated for such regimes. Particle filters rely on a Monte Carlo method of approximating the posterior distribution of an underwater moving target's location by a weighted set of sample points (particles) that evolve in time with the target. Variational Bayesian inference uses a Monte Carlo approach to estimate a distribution of the parameters of the posterior through a series of converging approximations. Our VB approach is shown to outperform the standard PF in both an active system (target tracking) and a passive system (source localization) in simulations.

Session 2aUW

Underwater Acoustics and Acoustical Oceanography: Mid- to High-Frequency Propagation and Scattering with Application to Underwater Communications

Mohsen Badiey, Cochair
College of Marine and Earth Sciences, Univ. of Delaware, Newark, DE 19716-3501

Daniel Rouseff, Cochair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105-6698

Chair's Introduction—8:00

Invited Papers

8:05

2aUW1. The dependence of the intensity of surface scattered signals on surface wave characteristics. James Preisig (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543, jpreisig@whoi.edu), Grant Deane (Univ. of California, La Jolla, CA 92093), and Chris Tindle (Univ. of Auckland, Auckland, New Zealand)

The compensation for the fluctuations in surface scattered signals remains one of the challenges for reliable and high rate underwater acoustic communications systems. Communication using phase coherent signals relies on the ability of the receiver to track the fluctuations in the impulse response of the channel between the transmitter and receiver. Techniques that can exploit *a priori* statistical characterizations of the fluctuations of the arrivals in the channel impulse response (fading statistics) will generally outperform systems that do not have access to such information. While fading statistics in radio communications is generally assumed to be Rayleigh or Rician, prior work [J. C. Preisig and G. B. Deane, J. Acoust. Soc. Am. **116**, 2067–2080 (2004)] has shown that surface wave focusing can significantly change channel fading statistics. The physical basis for the intensity of surface scattered signals for both isolated specular scattering points and caustics developed in [C. T. Tindle and G. B. Deane, J. Acoust. Soc. Am. **117**, 2783–2794 (2005)] relates intensity to ray geometry parameters. In this work, surface wave characteristics are related to both the source to receiver ranges at which normal focused specular scattering and caustics will occur and to the intensities of the surface scattered arrivals in each case. Experimental data are presented comparing theoretically predicted and measured intensities.

8:25

2aUW2. Channel simulation for predicting acoustic modem performance. Michael B. Porter, Paul Hursky (HLS Res., 3366 N. Torrey Pines Ct., Ste. 310, La Jolla, CA 92037, mikeporter@hlsresearch.com), and Martin Siderius (Portland State Univ., Portland, OR 97207)

Acoustic modems usually provide the link in modern Distributed Netted Systems. However, their performance is strongly affected by the ocean weather as both the propagation medium and the source of background noise. Understanding these effects is important in pre-experiment planning, real-time adaptation, and retrospection (to ensure appropriate lessons are learned for future deployments). To this end, we have developed a channel simulator called VirTEX (Virtual Timeseries Experiment) that includes Doppler effects due to platform and ocean surface dynamics. VirTEX, in turn, is a component of the AComms Advisor, which provides links to ocean forecasts and other environmental information and uses HIL (Hardware in the Loop) testing to include a faithful modem simulation. This talk will discuss the process, as well as progress, in the algorithms.

8:45

2aUW3. Synthetic aperture communications exploiting fluctuations of the ocean environment. H. C. Song, W. S. Hodgkiss, W. A. Kuperman, T. Akal (Marine Physical Lab., Scripps Inst. of Oceanogr., UC San Diego, La Jolla, CA 92093-0238), and M. Stevenson (Spawar Systems Ctr., San Diego, CA 92152-5001)

Typically, fluctuations of the ocean environment pose difficulties for high-rate underwater acoustic communications. On the other hand, these fluctuations can be exploited in synthetic aperture communications (SAC) that take advantage of the relative motion between a transmitter and a receiver to create the equivalent of an array aperture. SAC requires that the channel impulse responses are sufficiently different from each other between transmissions for diversity combining to be effective. Rapid channel variations facilitate a shorter interval between transmissions resulting in a shorter integration time to achieve the desired diversity. This paper demonstrates high-rate SAC in shallow water exploiting fluctuations of the ocean environment.

9:05

2aUW4. Coherent communications performance in environments with strong multi-pathing: Model/data comparison. Daniel Rouseff (Appl. Phys. Lab., College of Ocean and Fishery Sci., Univ. of Washington, Seattle, WA 98105)

Coherent underwater communications is hampered by the time spread caused by multi-pathing and the Doppler spread caused by the rapidly changing shallow water environment. In practice, a time-varying impulse response model will retain only a finite number of paths as useful signal and will be updated to compensate for the changing environment at a finite rate. In the present study, communications performance is examined as a function of the number of retained paths and the model-update interval. Results are quantified in terms of the mean squared error in the soft demodulation output. A physics-based performance model is developed. The model uses as input acoustic quantities like the Rayleigh parameter and environmental quantities like the significant wave height and the dominant period for surface waves. Model predictions for performance are compared to experimental results for data collected near the Hawaiian Island of Kauai. [Work supported by ONR.]

9:20

2aUW5. Ocean variability effects on high-frequency coherent acoustic communications. Aijun Song, Mohsen Badiey (Robinson Hall 114C, College of Marine and Earth Studies, Univ. of Delaware, Newark, DE 19716,), H. C. Song, and William Hodgkiss (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

The variability of the ocean environment can cause fluctuations of acoustic channels and these fluctuations often pose serious difficulties for high rate digital communications. During a series of recent acoustic communications experiments conducted around Kauai Island, Hawaii (KauaiEx'03, MakaiEx'05, and KAM'08), coherent communication data with concurrent environmental measurements were collected under different experimental settings. The variability of the ocean parameters, mainly the water temperature profiles and the sea surface condition, has been studied in order to understand their impact on the performance of acoustic communications at high frequencies (greater than 10 kHz). It is shown that both the long-term (in hours) changes in the ocean volume and the fast fluctuations (in seconds) of the sea surface have impact on the receiver design and on the receiver performance. Using time reversal combining followed by a single channel decision feedback equalizer, the ocean variability effects on the performance of coherent acoustic communications are shown as a function of receiver depth, source depth, and communication range.

9:35

2aUW6. Sonar simulation toolset modeling for underwater acoustic communications and networking. Julia B. Hsieh (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698, julia@apl.washington.edu), Daniel Rouseff (Univ. of Washington, Seattle, WA 98105-6698), Aijun Song, and Mohsen Badiey (Univ. of Delaware, Newark, DE 19716)

The sonar simulation toolset (SST) is a ray-based propagation model capable of generating time series realizations with stochastic spreading of acoustic sequences. The model supports varying environments and geometries including moving sources and/or receivers. The simulated data were compared with experimental data collected at the Pacific Missile Range Facility, Kauai, Hawaii in 2003 (KauaiEx2003) and 2008 (KAM08). Models generated with appropriate sound speed profiles and wind speeds are shown to have similar channel impulse response functions to those observed during the experiments. The high-frequency communications sequences simulated from those models have been demodulated to demonstrate analogous results. These results suggest that underwater acoustic communications and networking can be effectively simulated with sonar simulation toolset to aid in experiment planning and data analysis. [Work supported by the Office of Naval Research.]

9:50

2aUW7. Basic research on time-reversal communication with a moving source-receiver. Takuya Shimura, Yoshitaka Watanabe, Hiroshi Ochi (Adv. Marine Tech. R&D 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)

A basic research on time reversal, to be used for communication with a long cruising AUV, has been researched by our research group. In such application, it is preferred to maintain its cruising speed. However, time reversal does not compensate the effect of moving source-receiver in principle. So in this study, simulations with a normal mode method considering the effect of source-receiver movement are executed in which the carrier frequency is 500 Hz and symbol rate is 100 bps. The method of combining time reversal and adaptive decision feedback equalizer is applied. In both cases of shallow water and the deep ocean, demodulation cannot be realized with only time reversal as expected, while the proposed method achieves communication at the moving speed up to 1.0 m/s.

10:05—10:20 Break

10:20

2aUW8. Impact of bottom type on orthogonal frequency division multiplexing in underwater communications. Nathan Parrish, Sumit Roy (Dept. of Elec. Eng., Univ. of Washington, Seattle, WA 98195-2500, nparrish@u.washington.edu), and Payman Arabshahi (Univ. of Washington, Seattle, WA 98105-6698)

The impact of bottom sediment type in relation to acoustic communications via orthogonal frequency division multiplexing (OFDM) is shown via experimental results and simulation. Experimental data from Lake Washington, Seattle with a "silty clay" bottom show that the multipath delay spread is longer at 250 m than at 4 km. This results in better OFDM performance at the longer range. Similar results are shown via simulation using a channel model developed from Bellhop, a Gaussian Ray tracing tool [M. Porter, "Bellhop Gaussian beam/finite element beam code," Available: http://oalib.hlsresearch.com/Rays/index.html (2007)]. Through simulation, results are also shown under similar conditions to the experiment but with varying bottom type. The results show that the performance of OFDM signaling is dependent on the bottom type as well as specific source/receiver geometry. [Work supported by NASA ESTO.]

10:35

2aUW9. Capacity of orthogonal frequency division multiplexing systems over fading underwater acoustic channels. Chantri Polprasert, James A. Ritcey (Dept. of Elect. Eng., Univ. of Washington, Seattle, WA 98195-2500, ritcey@ee.washington.edu), and Milica Stojanovic (Northeastern Univ., Boston, MA 02115)

The channel capacity bounds of orthogonal frequency division multiplexing (OFDM) systems over the underwater (UW) acoustic fading channel as a function of the distance between the transmitter and receiver are investigated. The upper bound is obtained under perfect channel state information (CSI) assumption at the receiver. The lower bound is obtained under imperfect CSI at the receiver. The reduction from the upper bound is due to the linear minimum mean square prediction error. The UW channel deviates from the wide sense stationary and uncorrelated scattering (WSSUS) model commonly used for small bandwidths. Frequency-selectivity is incorporated due to the acoustic propagation into each arrival path between the transmitter and receiver. This leads the UW channel to be modeled as a frequency-dependent doubly spread fading channel characterized by the wide sense stationary and correlated scattering (WSS-non-US) fading assumption. Both Rayleigh and Ricean fading assumptions are investigated in the model. Results from the model show a gap between the upper and lower bounds that depends not only on the ranges and shape of the scattering function of the UW channel but also the distance between the transmitter and receiver. The model for the scattering function was confirmed from experimental data.

10:50

2aUW10. Multichannel combination investigations for differential frequency hopping transmissions in shallow water. Dianne E. Egnor, Geoffrey S. Edelson (BAE Systems, Advvaned Systems & Technol., P.O. Box 868, Nashua, NH 03061-0868), Luca Cazzanti, and Julia Hsieh (Univ. of Washington, Seattle, WA 98105)

Underwater acoustic communication requires waveforms that are robust to the signal distortions caused by the acoustic channel. Many waveforms used for this purpose require the transmission of training symbols that span the intersymbol interference to "learn" and compensate for these channel effects. These waveforms also require exacting coordination between the transmitters to avoid multiple access interference. Differential frequency hopping (DFH) is a fast frequency hopping, digital signaling technology that requires minimal information at the transmitter to communicate in the underwater channel. DFH has the desirable performance features of noninterfering spread spectrum operation, spectral reuse, and fading and interference resistance. This presentation describes the baseline autosynchronizing, single-user DFH decoder for a multiple hydrophone receiver and investigates two processing techniques incorporated for shallow-water multiuser applications: fading mitigation and multiuser interference mitigation, as implemented in conjunction with array processing. We present the performance of the baseline DFH decoder in terms of bit error rate with and without these enhancements on single- and multiuser data collected at sea during the 2008 Surface Processes and Acoustic Communications Experiment (SPACE08).

11:05

2aUW11. Modeling differential frequency hopping communication in the underwater acoustic channel. Luca Cazzanti, Julia Hsieh (Appl. Phys. Lab., Univ. of Washington, Box 355640, Seattle, WA 98105), Dianne Egnor, and Geoffry S. Edelson (BAE Systems, Nashua, NH 03061)

Differential frequency hopping (DFH) is a fast frequency hopping digital modulation scheme with proven multiple-access and jamming robustness properties in typical wireless channels. Characterizing the capabilities of DFH modulation in the more challenging underwater acoustic channel requires careful analyses that rely on both computer simulations and data collected at sea. The Sonar Simulation Toolkit (SST) is used to model challenging underwater environments and simulate the propagation of DFH waveforms in the corresponding underwater acoustic channels. Our simulations provide baseline performance results that can be used to assess the capabilities of DFH modulation and guide future algorithmic improvements to the receiver. In particular, our simulations show that incorporating equalization techniques into the DFH baseline receiver leads to improved decoding performance in challenging environments characterized by long channel impulse responses, which are known to cause inter-symbol interference in the received acoustic bit stream. To ensure their relevance to practical applications, our computer-based SST models are calibrated to the environmental parameters measured during recent at-sea experiments, including the Rescheduled Acoustic Communications Experiment (RACE08), and the corresponding performance is analyzed.

11:20

2aUW12. Achieving maximum space-time-frequency diversity in shallow water acoustic channels: System design and experimental results. Costas Pelekanakis and Arthur B. Baggeroer (5-206, Ctr. for Ocean Eng., Dept. of Mech. Eng., Massachusetts Inst. of Technol., 77 Massachusetts Ave., Cambridge, MA 02139, gas@mit.edu)

This work sheds light on the individual impact of coded modulation and diversity in achieving reliable underwater acoustic (UWA) communications. Two systems with the same bit-rate and complexity but different diversity order are designed and compared. The "conventional" system combines trellis coded modulation (TCM), symbol interleaving and an 8PSK signal set. The "alternative" or bit-interleaved coded modulation (BICM) system combines a convolutional code, bit interleaving, and a 16QAM signal set. Both systems are coupled with orthogonal frequency division multiplexing (OFDM) to cope with the intersymbol interference (ISI). When two or three projectors are used, the above coding schemes are concatenated with space-time block codes (STBC). Bit-error rates using data recorded during the RACE'08 off the coast of Narragansett, USA are computed demonstrating the superiority of BICM schemes relative to their TCM counterparts over an extensive collection of shallow water channels, thus proving that BICM is a promising technique towards achieving reliable UWA communications. In particular, by using BICM and 4 kHz of bandwidth, we have successfully accomplished a 7150 b/s transmission rate over 1 km range.

11:35

2aUW13. Improving underwater vehicle communication in the littoral zone through adaptive vehicle motion. Toby E. Schneider and Henrik Schmidt (Ctr. for Ocean Eng., Dept. of Mech. Eng., MIT, 77 Massachusetts Ave., Cambridge, MA 02139, tes@mit.edu)

Autonomous underwater vehicles (AUVs) are increasingly being used in clusters to perform detecting and sensing tasks in the littoral zone. These tasks require substantial data throughput between underwater and surface nodes. Acoustic modems provide the only practical means of sending these data, but sound in the littoral waveguide is heavily refracted downward and multiply reflected by the shallow bottom. These effects lead to shadow zones where messages are lost and convergence zones that allow for longer range receipt of high bitrate messages. During the GLINT08 sea trials, messages were sent using a multiple rate 25-kHz acoustic modem (the WHOI "MicroModem") between several undersea and surface nodes. The message quality (mean square error) was analyzed and compared to dynamic ray tracing models of the transmission loss, using environmental data measured by the AUVs and the research vessel. Significant correlation was found between the data and the modeled acoustic field, suggesting that such models could be run on-board the AUVs to predict regions of low transmission loss. This would give the AUVs predictive ability to choose when to send messages to maximize the probability of receipt and the choice adaptively move to improve chances of successful message reception.

Meeting of Standards Committee Plenary Group

to be held jointly with the meetings of the
ANSI-Accredited U.S. Technical Advisory Groups (TAGs) for:
ISO/TC 43, Acoustics,
ISO/TC 43/SC 1, Noise,
ISO/TC 108, Mechanical vibration, shock and condition monitoring,
ISO/TC 108/SC 2, Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures,
ISO/TC 108/SC 3, Use and calibration of vibration and shock measuring instruments,
ISO/TC 108/SC 4, Human exposure to mechanical vibration and shock,
ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines,
ISO/TC 108/SC 6, Vibration and shock generating systems,
and
IEC/TC 29, Electroacoustics

P. D. Schomer, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 43 Acoustics and ISO/TC 43/SC 1 Noise
2117 Robert Dr., Champaign. IL 61821

D. J. Evans, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108 Mechanical vibration shock and condition monitoring, and ISO/TC 108/SC 3 Use and calibration of vibration and shock measuring devices
National Inst. of Standards and Technology, 100 Bureau Dr., Stop 8220, Gaithersburg, MD 20899

W. C. Foiles, Co-Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures
BP America, 501 Westlake Park Blvd., Houston, TX 77079

R. Taddeo, Co-Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures
NAVSEA, 1333 Isaac Hull Avenue, SE, Washington Navy Yard, Washington, DC 20376

D. D. Reynolds, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 4 Human exposure to mechanical vibration and shock
3939 Briar Crest Ct., Las Vegas, NV 89120

D. J. Vendittis, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 5 Condition monitoring and diagnostics of machines
701 Northeast Harbour Ter., Boca Raton, FL 33431

R. Taddeo, Vice Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 5 Condition monitoring and diagnostics of machines
NAVSEA, 1333 Isaac Hull Avenue, SE, Washington Navy Yard, Washington, DC 20376

C. Peterson, Chair
U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 6 Vibration and shock generating systems
200 Dixie Ave., Kalamazoo, MI 49001

V. Nedzelnitsky
U.S. Technical Advisor (TA) for IEC/TC 29 Electroacoustics
National Inst. of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD 20899–8221

The reports of the Chairs of these TAGs will not be presented at any other S Committee meetings.

The meeting of the Standards Committee Plenary Group will precede the meetings of the Accredited Standards Committees S1, S3, S3/SC 1, and S12 which are scheduled to take place in the following sequence:

ASC S3, Bioacoustics	**Tuesday, May 19**	**10:30 a.m. - 11:30 a.m.**
ASC S3/SC 1, Animal Bioacoustics	**Tuesday, May 19**	**11:45 a.m. - 1:00 p.m.**
ASC S1, Acoustics	**Tuesday, May 19**	**2:45 p.m. - 3:45 p.m.**
ASC S12, Noise	**Tuesday, May 19**	**4:00 p.m. - 5:00 p.m.**

Please note that ASC S2, Mechanical Vibration and Shock will meet in advance of the Plenary meeting, on Monday, May 18 at 4:30 p.m.

Discussion at the Standards Committee Plenary Group meeting will consist of national items relevant to all S Committees and U.S. TAGs.

The U.S. Technical Advisory Group (TAG) Chairs for the various international Technical Committees and Subcommittees under ISO and IEC, which are parallel to S1, S2, S3, and S12 are as follows:

U.S. TAG Chair/Vice Chair	TC or SC	U.S. Parallel Committee
ISO		
P. D. Schomer, Chair	**ISO/TC 43** Acoustics	S1/S3
P. D. Schomer, Chair	**ISO/TC 43/SC1** Noise	S12
D. J. Evans, Chair	**ISO/TC 108** Mechanical vibration, shock and condition monitoring	S2
W. C. Foiles, Co-Chair R. Taddeo, Co-Chair	**ISO/TC 108/SC2** Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures	S2
D. J. Evans, Chair	**ISO/TC 108/SC3** Use and calibration of vibration and shock measuring instruments	S2
D. D. Reynolds, Chair	**ISO/TC 108/SC4** Human exposure to mechanical vibration and shock	S2/S3
D. J. Vendittis, Chair R. Taddeo, Vice Chair	**ISO/TC 108/SC5** Condition monitoring and diagnostics of machines	S2
C. Peterson, Chair	**ISO/TC 108/SC6** Vibration and shock generating systems	S2
IEC		
V. Nedzelnitsky, U.S. TA	**IEC/TC 29** Electroacoustics	S1/S3

Meeting of Accredited Standards Committee (ASC) S3 Bioacoustics

C. A. Champlin, Chair S3
Univ. of Texas, Dept. of Communication Sciences & Disorders, CMA 2–200, Austin, TX 78712

D. A. Preves, Vice Chair S3
Starkey Laboratories, Inc., 6600 Washington Ave., S. Eden Prairie, MN 55344

Accredited Standards Committee S3 on Bioacoustics. Working group chairs will report on the status of standards under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAGs for ISO/TC 43 Acoustics and IEC/TC 29 Electroacoustics, take note — those meetings will be held in conjunction with the Standards Plenary meeting at 9:00 a.m. on Tuesday, 19 May 2009.

Scope of S3: Standards, specifications, methods of measurement and test, and terminology in the fields of psychological and physiological acoustics, including aspects of general acoustics which pertain to biological safety, tolerance and comfort.

Meeting of Accredited Standards Committee (ASC) S3/SC 1 Animal Bioacoustics

D. K. Delaney, Chair S3/SC 1
USA CERL, 2092 Farber Dr., Champaign, IL 61822

M. C. Hastings, Vice Chair S3
6422 Crosswoods Dr., Falls Church, VA 22044 12144

Accredited Standards Committee S3/SC 1 on Animal Bioacoustics. Working group chairs will report on the status of standards under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

Scope of S3/SC 1: Standards, specifications, methods of measurement and test, instrumentation and terminology in the field of psychological and physiological acoustics, including aspects of general acoustics which pertain to biological safety, tolerance and comfort of non-human animals, including both risk to individual animals and to the long-term viability of populations. Animals to be covered may potentially include commercially-grown food animals; animals harvested for food in the wild; pets; laboratory animals; exotic species in zoos, oceanaria or aquariums; or free-ranging wild animals.

Session 2pAA

Architectural Acoustics: Computer Auralization

Robert C. Coffeen, Chair
School of Architecture and Urban Design, Univ. of Kansas, Lawrence, KS 66045

Chair's Introduction—1:00

Invited Papers

1:05

2pAA1. Auralization–audibility of changing scattering coefficients for audience areas in different size venues. Sarah Rollins (Sparling, Inc., 720 Olive Way, Ste. 1400, Seattle, WA 98101, srollins@sparling.com)

Auralizations in computer modeling are affected by many different variables, such as absorption coefficients, scattering coefficients, and the complexity of the model. However, scattering coefficients are often determined by the best guess of the consultant. In order to isolate the changes due to the scattering coefficient for comparison with measured data, a simple rectilinear room was measured and modeled. This paper will show how changing the architectural surface and audience scattering coefficients affects the auralizations for this room. For comparison, the effects of changing these same coefficients in a larger venue will also be presented.

1:25

2pAA2. Comparison of modeled versus measured room acoustic parameters in the recently completed Shaghoian Concert Hall in Clovis, California. Bill Dohn (Dohn and Assoc., Inc., 630 Quintana Rd. 312, Morro Bay, CA 93442, bill.dohn@gte.net) and Richard H. Campbell (Bang-Campbell Assoc., Falmouth, MA 02536)

The Shaghoian Concert Hall in Clovis, CA is a popular new variable acoustic music venue in the greater Fresno, CA area. The design process for this hall included CATT-Acoustic (TM) modeling for analysis and verification of essential room acoustic parameters based upon the settings of the variable acoustic devices. The model of the final design produced parameters that were closely in line with ideal values for a hall of this size and use. Model results will be compared to measured parameters in the completed hall.

1:45

2pAA3. From toy to tool—advanced auralization procedure in daily design work. Wolfgang Ahnert and Stefan Feistel (AFMG, Arkonastr. 45-49, 13189 Berlin, Germany)

Since more than 20 years simulation procedures are more and more common, but the auralization procedure is considered mainly as a toy. The reason for this relates partially to the efforts needed to get real auralized sound files. In a previous paper the authors reported about a method to produce high end auralization files. This method allows to use normal head phones to get the right spatial impression without in-head localization. The main reason not to use an auralized sound impression in the daily design work is the time-consuming procedure to calculate true impulse responses (IR) especially in large halls. In this paper new algorithms are reported to calculate IRs in multithread mode with one computer or better via a network with a set of CPUs working in parallel. So calculations in large halls maybe be reduced from 8–10 h with one CPU to 10–20 min with multicore computers. The paper explains the way from the model to high-end auralized files. Some examples demonstrate the quality of these derived binaural files.

2:05

2pAA4. The spatial perception of a large pipe organ in various locations in a church computer model is investigated using multiple sources properly spaced within an organ chamber, with each source representing an organ stop. Richard Campbell (Bang-Campbell Assoc., 26 G Chilmartk Dr. East Falmouth, MA 02536)

It is surmized that the auralization of a large pipe organ should use multiple sources properly spaced in the organ chamber with each source representing a stop. But at some point in the acoustic environment of the listener, the advantage of such spatial sourcing disappears. A MIDI sequencer is used to generate audio files of a Buxtahuda fugue in six parts using six different stops. The sources are properly spaced within an organ chamber in a large church. Binaural recordings are evaluated at various locations in the church.

2:25

2pAA5. Comparison of auralized and measured sound in moderately reverberant spaces. Robert C. Coffeen, Jonathan Birney, Stephanie Hoeman, Shane Kanter, Hannah Schultheis (School of Architecture and Urban Planning, Univ. of Kansas, 1465 Jayhawk Blvd., Lawrence, KS 66045, coffeen@ku.edu), Lauren M. Ronsse, and Lily M. Wang (Univ. of Nebraska-Lincoln, Omaha, NE 68182-0681)

This research performs an in-depth comparison of sound from computer auralizations with measured binaural room impulse responses and recordings taken in actual spaces. Binaural measurements and recordings have been gathered in two church naves with midfrequency reverberation times of about 1.5 and 2.5 s. The results from these measurements are compared with auralizations from computer models of the spaces in two different room acoustic modeling programs. The source and receiver positions used for the auralizations are the same as those measured in the real spaces. One of the naves was previously investigated [R.C. Coffeen and G. Caunt, J. Acoust. Soc. Am. **105**, 1174 (1999) (A)] with results from computer auralizations compared to measured binaural recordings in the space. As numerous advancements have been made to computer modeling programs since that time, this project revisits and expands the previous research using updated modeling and measurement techniques. The results of this research not only portray the developments in computer modeling that have occurred over the past decade, but also depict the relationship between current computer auralizations and actual sound in moderately reverberant spaces.

2:55

2pAA6. Using computer auralization to help prevent major problems relating to room acoustics and electroacoustic sound reinforcement. Robert C. Coffeen (Sch. of Archit. and Urban Plan., Univ. of Kansas, 1465 Jayhawk Blvd., Lawrence, KS 66045, coffeen@ku.edu)

Computer modeling and computer auralization will produce comparative listening experiences for architects, building owners, and building users based on selected interior surface shapes, interior finish materials, and electroacoustic sound distribution. Auralization can be a significant tool in helping building designers, owners, and users understand potential acoustical problems and the need for proper acoustic and electroacoustic design.

Contributed Papers

3:15

2pAA7. Diffraction culling for virtual-acoustic simulations. Paul Calamia (Dept. of Comp. Sci., Princeton Univ., 35 Olden St., Princeton, NJ 08540, pcalamia@cs.princeton.edu), Benjamin Markham (Acentech Inc., Cambridge, MA 02138), and U. Peter Svensson (Norwegian Univ. of Sci. and Technol., NO-7491 Trondheim, Norway)

Acoustic simulations of complex virtual environments typically are created with geometrical-acoustics techniques. Such simulations can be augmented with edge diffraction modeling for improved accuracy, but not without a significant increase in processing time due to the additional propagation paths which must be considered and the computational complexity of the diffraction calculations. However, for a given modeling scenario, the contribution of a diffracted path to the overall impulse response can vary over a large range, suggesting that certain diffracted paths can be ignored, or culled, to reduce processing time with a limited effect on the accuracy of the simulation. In this talk, we first analyze the effects of diffraction culling through a precomputed, amplitude-based ranking scheme. We then describe a simple procedure for identifying and culling insignificant diffraction components during a virtual-acoustic simulation which approximates the performance of the precomputed ranking. Through numerical and subjective analysis, we show that a significant percentage of diffracted paths can be ignored if the retained paths are those which lead to the highest-amplitude diffraction components, although the audible effects of such diffraction culling are dependent on the input signal.

3:30

2pAA8. Singers' preferences for acoustical characteristics of performing spaces. Kathleen Stetson (Arup Acoust., 155 Ave. of the Americas, New York, NY 10013, kathleen.stetson@arup.com) and Jonas Braasch (Rensselaer Polytechnic Inst., Troy, NY 12180)

Classical singers, whose instruments' close proximity to their ears makes them unique among musicians, require particular attention from acousticians addressing musician support in performance space design. This study expands upon the few previous analyses of acoustics for singers by exploring what is most basic about solo voice self-perception in halls used for concerts and recitals. A questionnaire was given to a number of professional classical singers and their numerical and narrative responses were analyzed. Five concert halls were measured from the singer's perspective on stage, utilizing a head and torso simulator with a mouth speaker to mimic the proximity of a singer's voice and ears. The resulting impulse responses were utilized in preference tests featuring real-time binaural auralization of singers' vocalizations. The correlations between the subjective results, existing objective parameters, and physical hall characteristics were explored. A strong connection was found between increasing preference and increasing reverberation time. Additionally, test subjects indicated a statistically significant dislike of or preference for a hall, regardless of the singing location on stage or the classical genre being sung. Further developments of this study are discussed in the context of ongoing opera house design utilizing computer model simulations.

3:45

2pAA9. Thirteen findings in ten minutes. Michael Ermann, Nate Crawford, Braden Field, Jessica Green, Sky Kim, Julia Mitchell, Steve Smith, Matthew Van Wagner, John Samuel Victor, and Amanda Massengill (School of Architecture and Design, Virginia Tech, 201 Cowgill Hall, Blacksburg, VA 24061-0205, mermann@vt.edu)

Nine architecture students in an architectural acoustics course studied the influence of design on the acoustic response of spaces. Thirteen of their projects will be presented, rapidly, with time to delve into some of them more deeply in response to questions from attendees. Areas of focus include: room shaping and balcony composition; ceiling and canopy height; canopy design; perception and preference of double sloped decays; side wall angle and orientation; room length; very tall ceilings; room width; side wall diffusion; mapping background noise levels; and frequency content of noise.

Session 2pAB

Animal Bioacoustics and Acoustical Oceanography: Autonomous Remote Monitoring Systems for Marine Animals II

Kathleen C. Stafford, Chair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105

Invited Papers

1:00

2pAB1. Validating acoustic monitors for marine animals: Field experience with beaked whales and digital acoustic monitors (DMONs). Mark Johnson, Tom Hurst (Woods Hole Oceanograph. Inst., 86 Water St., Woods Hole, MA 02543, majohnson@whoi.edu), Anton Arias, and Natacha Aguilar de Soto (Univ. of La Laguna, Tenerife, Spain)

Passive acoustic monitoring (PAM) is a promising tool for improving the effectiveness of visual surveys for marine animals and for stand-alone persistent monitoring. Nonetheless, significant data gaps make it difficult to predict PAM performance for many species. Detection functions are only beginning to be formed for some vocalizations and detectors. Statistical models for calling rates, as a function of behavior, are needed to translate acoustic detections into abundance estimates, or lack of detections into decisions about absence of animals. Finally, ambient noise levels and spectra are unavailable in most of the world's oceans, complicating the application of results from one area to another. Although some of these gaps are being addressed by a growing group of researchers, there is an urgent need to develop standards for measuring and reporting results, and to foster the interchange of data. To exemplify the form such standards might take, and the potential pitfalls, we present results from a series of field experiments designed to assess the performance of an acoustic monitor for beaked whales. The experiments made use of a freely available reference design for an acoustic detector, called the DMON, which may provide a basis for standardizing and validating real-time implementations.

1:20

2pAB2. Screening large data sets and real-time data streams for bioacoustic signals. Holger Klinck (Cooperative Inst. for Marine Resour. Stud., and NOAA Pacific Marine Environ. Lab., Oregon State Univ., 2030 SE Marine Sci. Dr., Newport, OR 97365, Holger.Klinck@oregonstate.edu), Lars Kindermann (Alfred Wegener Inst. for Polar and Marine Res., 27568 Bremerhaven, Germany), David K. Mellinger (Oregon State Univ., Newport, OR 97365), and Olaf Boebel (Alfred Wegener Inst. for Polar and Marine Res., 27568 Bremerhaven, Germany)

This presentation will discuss a major challenge of passive-acoustic monitoring systems: the analysis of large data sets to identify the occurrence of bioacoustic signals of interest. Triggered by the rapid development of digital audio technology and the increasing capacity of memory devices, it has become easier than ever to produce large, long-term acoustic data sets. However, analyzing these data sets is challenging, as suitable automated detection and classification systems are needed to perform the analysis in a reasonable amount of time. One such large data set is currently produced by the Perennial Acoustic Observatory in the Antarctic Ocean. This passive acoustic observatory has been operated by the Alfred Wegener Institute since January 2006 and has generated more than 10,000 hours of data to date. The observatory features a data link via satellite, which allows analysis of the recorded hydroacoustic data in real time in Germany. However, to be able to run custom algorithms over the entire data set in several times real-time speed, a distributed computing system was developed and applied. Here we provide a detailed description of this system and discuss further possible applications.

1:40

2pAB3. Acoustic sampling for marine mammals in the Beaufort Sea July 2007–March 2008. Kathleen Stafford (Appl. Phys. Lab, Univ. of Washington, 1013 NE 40th St., Seattle WA 98105, stafford@apl.washington.edu), Sue Moore (Alaska Fisheries Sci. Ctr., Seattle WA 98115), Catherine Berchok (Natl. Marine Mammal Lab, Seattle WA 98115), and David K. Mellinger (Oregon State Univ., Newport OR 97365)

As climate change is driving rapid, unprecedented warming of the Arctic, there is increasing interest in how such change will impact Arctic marine mammals. Impacts are anticipated from from habitat alteration, including increasing ambient noise levels from shipping, seismic exploration for oil and gas and geophysical research, and (potentially) commercial fishing. In order to monitor natural and anthropogenic sources of noise, four autonomous recorders were deployed along the 100-m isobath between Cape Halkett and Barrow and recorded data from July 2007–March 2008. The instruments sampled at 8192 Hz on a schedule of 10 min on, 20 min off. Marine mammal sounds recorded included pinnipeds (walrus and bearded seals) and cetaceans (bowhead and beluga whales), while anthropogenic sources included shipping and air gun sounds. Seasonal and geographic patterns for these sounds will be presented. These data form part of a broader-scale international, year-round monitoring program in the Arctic that we hope will eventually span the entire Arctic and provide a basin-wide acoustic observatory.

2:00

2pAB4. Integration of automated detection methods into NOAA Southwest Fisheries Science Center (SWFSC) acoustic marine mammal monitoring protocol. Tina M. Yack, Jay Barlow, Shannon Rankin (Natl. Marine Fisheries Svce., Southwest Fisheries Sci. Ctr., 3333 N. Torrey Pines Court, La Jolla, CA 92037, tina.yack@noaa.gov), and Douglas Gillespie (Univ. of St. Andrew, Fife KY16 8LB, Scotland)

Southwest Fisheries Science Center (SWFSC) has used combined visual and acoustic techniques to monitor marine mammal populations for the past 8 yrs. Currently, SWFSC passive acoustic surveys of cetaceans require specially trained personnel to monitor hydrophone signals in real-time. While effective, this method is time-consuming and costly. Automated detection of cetacean vocalizations would be a valuable tool during SWFSC surveys, allowing for detection when experienced technicians are unavailable. This technique is advantageous because it significantly reduces effort and removes sources of human error and bias in detection ability. PAMGUARD 1.0 CORE software was evaluated for use in automated detection of cetacean acoustic signals. Three different detector configurations of PAMGUARD were evaluated. This work shows that the majority of whistle and click events can be detected using PAMGUARD software. All of the PAMGUARD trials were capable of detecting whistles and clicks of cetacean species with varying success. These techniques were field-tested at sea during a recent SWFSC marine mammal survey. Automated detection of beaked whales and Dall's porpoise during this survey will also be discussed. It is our goal to integrate automated detection methods into SWFSC's acoustic marine mammal monitoring protocol and this work is an important step in doing so.

2:15

2pAB5. Tracking fin and blue whales above the Juan de Fuca Ridge with a local seafloor seismic network. William S. D. Wilcock, Dax C. Soule (School of Oceanogr., Univ. of Washington, Box 357940, Seattle, WA 98195, wilcock@u.washington.edu), and Richard E. Thomson (Inst. of Ocean Sci., Sidney, BC, V8L 4B2 Canada)

The Endeavour segment of the Juan de Fuca mid-ocean ridge hosts several high-temperature hydrothermal fields. Previous analysis of bioacoustical data shows that zooplankton are enhanced at all depths above the hydrothermal vent fields compared with sites $\geqslant$10 km away. From 2003–2006, a seafloor seismic network was deployed around the hydrothermal vent fields to monitor earthquakes and it also recorded an extensive data set of fin and blue whale calls. As part of an investigation of a potential correlation between whale tracks, enhanced zooplankton concentrations, and hydrothermal vents above the Juan de Fuca Ridge, an automatic algorithm is being developed to track vocalizing whales that swim near the network. Events are detected by triggering with the ratio of short-term to long-term running RMS averages and whale calls are distinguished from earthquakes based on their spectra. For fin whales each 1-s arrival is identified based on its instantaneous amplitude and frequency and a pick is made at the mid-energy point. A grid search method is used to localize calls using direct and multipath arrivals. The algorithm and preliminary results will be presented. [The Keck Foundation supported the seismic network and the Office of Naval Research is supporting this study.]

2:30

2pAB6. Frequency quantiles and dual harmonic tracking for detection and classification of killer whale calls. Val Veirs (The Whale Museum, Friday Harbor, WA 98250, Colorado College, Colorado Springs, CO, val@beamreach.org)

The Salish Sea Hydrophone Network, www.OrcaSound.net, is operating five streaming hydrophones in the critical habitat of the endangered Southern Resident orcas. Software is being developed that allows each site to sift through the approximately 10 Gbytes of information that are produced each day at each site in order to detect killer whale calls and related sounds. Detection of calls is accomplished via triggering on three time dependent features extracted in real-time from sequential power spectra. One of these three features is derived from the bandwidth between quartiles of the filtered power spectrum. The other two quantify the degree to which the power spectrum has one or two harmonic structures. Features are extracted into feature sets that save the details of each feature: amplitudes; durations; and time variations. A metric has been devised that measures the similarity of different sequences of feature sets. With this metric an unsupervised learning algorithm builds a tree structure that classifies sounds by their similarity. As new calls come in, they either fall into pre-existing groups or create new branches. Automatic reports are made via the Web and e-mail.

2:45

2pAB7. Tracking multiple sperm whales with widely spaced bottom-mounted hydrophones. Eva-Marie Nosal (Dept. of Ocean and Resources Eng., Univ. of Hawaii, 2540 Dole St., Honolulu, HI 96816)

One dataset from the 2nd Int. Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics Data featured multiple clicking sperm whales recorded for 20 min. Data were recorded on five widely spaced bottom-mounted hydrophones at the Atlantic Undersea Test and Evaluation Center. This dataset is used to develop and apply a model-based time difference of arrival (TDOA) method that is capable of simultaneously tracking multiple animals. Clicks on all hydrophones are detected. For every receiver pair, all possible time-differences of arrival (within reasonable time limits) are formed from the detected clicks. A scatterplot of TDOAs versus time, although noisy due to many false TDOAs formed from incorrectly associated clicks, reveals clear "tracks" of slowly varying TDOAs. These tracks are extracted and input to a TDOA tracking algorithm that gives 3-D likelihood surfaces of whale position. Local maxima in the likelihood surfaces are extracted and clustered to give estimated whale tracks.

3:00—3:20 Break

Invited Paper

3:20

2pAB8. The Acousonde: A miniature autonomous wideband recorder. William C. Burgess (Greeneridge Sci. Inc., 6060 Graham Hill Rd., Fl. 2 Stop F, Felton, CA 95018, burgess@greeneridge.com)

The Acousonde™ is a newly designed miniature acoustic/ultrasonic recording tag incorporating several improvements over its predecessor, the Bioacoustic Probe. Design trade-offs for its acoustic data paths targeted acquisition of both near and distant odontocete echolocation clicks with minimal distortion while preserving general-purpose utility and low-power operation. Two acoustic channels are available: a low-power channel for long-term recording of signals up to 9 kHz and a high-frequency channel for signals up to 100 kHz. Each channel has its own dedicated hydrophone. For antialiasing, the low-power channel uses an adjustable switched-capacitor elliptic filter, while the high-frequency channel uses a fixed-frequency linear-phase filter; both filters may be bypassed if raw acquisition is desired. "Ping-pong" alternating sampling may be used to acquire samples from both channels concurrently, possibly to assess time-of-arrival differences between the two hydrophones. The controlling microprocessor, an ARM 9 with vector floating point accelerator, can digitally filter and downsample acoustic data during acquisition to reduce storage requirements. Other sensors include a 3D accelerometer, a 3D compass, a depth transducer, and a temperature monitor. Eight gigabytes of data storage are available, with data offload via a MicroUSB connector. Initial tests with captive animals are planned shortly. [Work supported by ONR.]

3:40

2pAB9. Validated reef fish sound scans of passive acoustic monitors on Hawaiian coral reefs. Timothy C. Tricas and Kelly Boyle (Dept. of Zoology and Hawaii Inst. of Marine Biol., Univ. of Hawaii at Manoa, Honolulu, HI)

Monitoring of fish sounds on coral reefs is a valuable potential technique for fisheries managers to remotely assess local fish populations and their behaviors. However, the species identity of most fish sounds on shallow coral reefs is not known. We have deployed two ecological acoustic recorders (EARs) at 20 m deep on two shallow reefs on the island of Hawaii. In order to validate which fish species produce these acoustic behaviors we used closed circuit rebreather diving, which produces almost no ambient acoustic noise, to record video and sonic behaviors of fish near the recording sites. From the videos we have identified 37 putative acoustic species many of which produce multiple sounds that include behaviors associated with intra- and interspecific aggression, feeding, courtship, and spawning. Cluster analysis was used to identify acoustic clades among and within species. Acoustic waveform average templates were constructed for different species clades and used in XBAT to screen the long-term recordings from the EAR. The use of these species-specific sound templates confirms the potential to detect diel, lunar, and longer-term rhythmicity for several species. This technique also shows promise for characterizing seasonal and annual periods of activity of sounds associated with context-specific behaviors.

3:55

2pAB10. Beaked whale density estimation from single hydrophones by means of propagation modeling. Elizabeth T. Küsel, David Mellinger (NOAA/PMEL, Hatfield Marine Sci. Ctr., Oregon State Univ., Newport, OR 97365, elizabeth.kusel@noaa.gov), Len Thomas, Tiago A. Marques (Univ. of St. Andrews, Scotland), David J. Moretti, and Jessica Ward (Naval Undersea Warfare Ctr., Newport, RI 02841)

Passive acoustic sonar systems offer many advantages to the study of marine mammals. For density estimation studies, it is important to evaluate the probability of detecting an animal as a function of its distance from the receiving sensor. In this work, acoustic propagation modeling is used to estimate the transmission loss as a function of depth and range between a source whale and a single-hydrophone receiver. The computed transmission loss is compared to ambient noise levels and source level distributions to estimate the detection probability as a function of range. Results will be compared to beaked whale data recorded on bottom-mounted sensors in the Atlantic Undersea Test and Evaluation Center (AUTEC) in the Bahamas, where the location of clicks is relative to one hydrophone. Source level and beam pattern extracted from digital acoustic tags (DTags) applied to a sample of animals at the same location will also be used in the detection model, and beaked whale spatial density will be estimated. The detection probability function will provide a relevant comparison to the detection function derived empirically from the DTag data by Marques *et al.* [Marques *et al.*, J. Acoust. Soc. Am. (submitted).]

4:10

2pAB11. Passive acoustic monitoring of fish activity in the Hawaiian Archipelago. Pollyanna Fisher-Pool, Marc O. Lammers (Joint Inst. for Marine and Atmospheric Res., NOAA-CRED, 1125B Ala Moana Blvd., Honolulu, HI 96814, Pollyanna.Fisher-Pool@noaa.gov), Kevin Wong, Russell E. Brainard (NOAA Fisheries, Honolulu, HI 96814), and Whitlow W. L. Au (Univ. of Hawaii, Kailua, HI 96734)

Fish are an important component of coral reef ecosystems and can be valuable indicators of ecological change on reef environments. Many species of fish produce sounds and are therefore well suited for passive acoustic monitoring. Since 2006, NOAA has been using Ecological Acoustic Recorders (EARs) to monitor coral reef habitats in the main and Northwestern Hawaiian Islands in an effort to develop multi-year time series of biological activity at these locations. Sixteen EARs are presently deployed on reef habitats ranging in depth from 2 to 23 meters. Preliminary results reveal that each site is characterized by a unique composition of acoustically active fish . Many sounds are also common at multiple sites. Different species show considerable variability in their diet and seasonal patterns of activity. The results indicate that long-term acoustic monitoring of fish at these sites is likely to yield useful information about changing patterns of fish presence, behavior, and relative abundance. These metrics can be used to gauge the relative stability of ecosystems and be related to physical, oceanographic, and anthropogenic variables affecting them.

4:25

2pAB12. Using moored passive acoustic recorders to assess seasonal occurrence and movements of southern resident killer whales in the coastal waters of Washington State. M. Bradley Hanson, Candice K. Emmons (NOAA/NMFS/Northwest Fisheries Sci. Ctr., 2725 Montlake Blvd. E, Seattle, WA 98112, brad.hanson@noaa.gov), Jeffrey A. Nystuen (Univ. of Washington, Seattle, WA 98105-6698), and Marc O. Lammers (Oceanwide Sci. Inst., Honolulu, HI 96839)

Designating critical habitat is mandated for species listed under the U.S. Endangered Species Act. This task has only been partly accomplished for southern resident killer whales (SRKWs) because winter distribution is poorly understood due to a variety of factors limiting visual sightings within their known central California to northern British Columbia range. To capitalize on the unique vocal behavior of resident killer whales, including pod-specific dialects, two types of acoustic recorders were deployed at strategic locations that span the Washington coast. Between 2005 and 2008 recorders were deployed in early winter for an average of 175.5 days at up to four sites. These functioned for an average of 114.8 days and collected a total of 47 SRKW detections. This exceeds the number of visual sightings during the same time period (15). Additionally, northern resident, transient, and offshore killer whales were recorded as well as Pacific white-sided dolphins, and humpback and sperm whales. SRKW were detected by both types of recorders and in all areas. Detections were made between Jan. and July with the majority of these detections in Mar., Apr., and May. This new information will be of key importance to managers in meeting recovery goals.

Session 2pAO

Acoustical Oceanography and Underwater Acoustics: Environmental Inferences in Inhomogeneous Ocean Environments I

David P. Knobles, Chair
Applied Research Labs., Univ. of Texas at Austin, Austin, TX 78713

Chair's Introduction—1:00

Invited Papers

1:05

2pAO1. Environmental influences on low-frequency, shallow-water acoustic propagation and inversion. George V. Frisk (Dept. of Ocean Eng., Florida Atlantic Univ., 101 N. Beach Rd., Dania Beach, FL 33004, gfrisk@seatech.fau.edu)

Historically, the seabed has been considered to play the dominant role in shallow-water acoustic propagation at low frequencies. As a result, propagation models have focused on the incorporation of accurate values of bottom properties, while inversion techniques have concentrated on the determination of geoacoustic properties of the seabed. In recent years, however, the assumption of a benign water column has increasingly come under scrutiny in addressing both the forward and inverse problems. This paper addresses these issues as they relate to a measurement and inversion methodology in which acoustic data are acquired on synthetic radial apertures that are created using a moving source/receiver configuration in concert with precision navigation. These data can be transformed into the horizontal wavenumber domain to obtain an estimate of the propagation characteristics of the waveguide; namely, its modal content. The modal spectra can then be inverted using a variety of techniques to determine estimates of the waveguide parameters, specifically the geoacoustic properties of the seabed. The issues associated with the successful implementation of this modal mapping method in a highly variable shallow-water environment are discussed. [Work supported by ONR.]

1:25

2pAO2. The impact of ocean sound speed variability on the uncertainty of geoacoustic parameter estimates. Ross Chapman and Yongmin Jiang (School of Earth and Ocean Sci. Univ. of Victoria, Victoria, BC V8W 5C2, Canada, chapman@uvic.ca)

This paper investigates the influence of water column variability on the estimates of geoacoustic model parameters obtained from matched field inversions. The acoustic data were collected on the New Jersey continental shelf during the Shallow Water 2006 experiments. The oceanographic variability was evident when the data were recorded. To quantify the uncertainties of the geoacoustic parameter estimates in this environment, Bayesian matched field geoacoustic inversion was applied to multi-tonal continuous wave data. The spatially and temporally varying water column sound speed was parametrized in terms of empirical orthogonal functions and included in the inversion. Its impact on the geometric and geoacoustic parameter estimates was then analyzed by the inter-parameter correlations. Two different approaches were used to obtain information about the variation of the water sound speed. One used only the profiles collected along the experimental track during the experiment, and the other also included observations from a larger area and a greater time period. The geoacoustic estimates from both the large and small sample sets were consistent. However, due to the diversity of the oceanic sound speed, more empirical orthogonal functions were needed in the inversion when more sound speed profile samples are used. [Work supported by ONR.]

1:45

2pAO3. Observed temporal statistics of acoustic travel time and intensity in the South China Sea. Ching-Sang Chiu and Christopher W. Miller, D. Benjamin Reeder, Justin M. Reeves, Steven R. Ramp (Dept. of Oceanography, Naval Postgraduate School, 833 Dyer Rd., Monterey, CA 93943, chiu@nps.edu)

The coupled ocean and acoustic variability in the Northeastern South China Sea basin was monitored using moored oceanographic and acoustic sensors during the 2005-2006 Windy Island Soliton Experiment (WISE). The temperature and current records captured prominent ocean variability at multiple scales, induced by mesoscale eddies, internal tides, and large amplitude internal waves. The latter two were transbasin, propagating from near the Luzon Strait, through the deep basin, onto the northeastern shelf. The concurrent acoustic measurements were attained from the transmission of a 400 Hz, phase-modulated signal along a 166 km path every 15 min. The receptions were processed to give the arrival structure of a pulse and its temporal change over a nine-month period. The observed temporal variability in the statistics of the acoustic travel time and intensity were analyzed using time-series techniques and models with emphasis to elucidate the connection and sensitivity to the observed ocean variability. Results from the analysis are discussed, as they pertain to the ocean processes and those parameters that can potentially be inferred from these types of acoustic transmissions. [The research is sponsored by the Office of Naval Research.]

2:05

2pAO4. Phase front fluctuations due to internal waves in shallow water 2006 experiment. Boris Katsnelson (Voronezh State Univ., Voronezh, Russia), Mohsen Badiey, Jing Luo (Univ. of Delaware, Newark, DE 19716), and James Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Experimental results and theoretical analysis of the sound phase front fluctuations measured by Vertical and Horizontal Line (VHLA) array (or L-shaped array) are presented. Low frequency broadband signals were radiated by the source (Miami Sound Machine) placed at the distance of ~25 km from the VHLA. During the time period 18:00–22:00 GMT on Aug. 17, 2006, a train of internal solitons traveled across the acoustic track and initiated fluctuations of the sound field. In an earlier paper [J. Acoust. Soc. Am. **124**, EL66–EL72 (2008)], it was shown that intensity of these fluctuations results from the horizontal refraction of the sound waves due to internal solitons. Here we show the acoustic phase front fluctuations for separate waveguide modes for the same period. To obtain phase fronts for the individual modes, a special approach using an L-shaped receiver array is used. The vertical part of the array was used to obtain modal amplitudes. Using the measured sound field based on the perturbation theory, modal phases at the HLA can be calculated. Results of data processing are compared with theoretical calculations. [Work supported by ONR, CRDF, and RFBR.]

2:20

2pAO5. Signal intensity fluctuations in Shallow Water 2006 (SW06) experiment during Event 50. Mohsen Badiey, Jing Luo (College of Marine and Earth Stud., Univ. of Delaware, Newark, DE 19716), Boris Katsnelson (Voronezh State Univ., Voronezh, Russia), and James Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

During the Shallow Water 2006 (SW06) experiment, several internal wave (IW) events were recorded by moving and stationary instruments simultaneously with the acoustic propagation tests. More than 50 IW events were recorded using the University of Delaware's R/V Sharp's radar. On this ship, a J15 sound source transmitted various acoustic signals at different bearings and ranges to the Woods Hole Oceanographic Institution's vertical and horizontal hydrophone arrays. During one of the IW events, the acoustic source was kept at a constant water depth while moving in horizontal plane with the advancing internal wave front. This internal wave packet, named Event 50, was also recorded by R/V oceanous radar a few kilometers away. The reason for placing the sound source on the advancing IW front was to examine the effects of azimuthal variability of the waveguide on the acoustic propagation. A detailed transition of the waveguide as the internal wave progressively occupied the acoustic propagation track between the source and receiver is reported. It is shown that the intensity peaks at small angles between the acoustic track and the internal wave front where sound intensity focusing and defocusing occurs. [Work supported by ONR 321OA and CRDF, and RFBR.]

2:35

2pAO6. Intensity measurements and 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), James F. Lynch, Aurthur E. Newhall (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), and Mohsen Badiey (Univ. of Delaware, Newark, DE 19716)

During the Shallow Water 2006 (SW06) experiment, the University of Delaware's R/V Sharp transmitted various acoustic signals at several different bearings and ranges to the Woods Hole Oceanographic Institute's vertical and horizontal hydrophone line array. The array was strategically positioned near the shelfbreak 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. The internal waves and the shelfbreak front can cause complex multimode and multipath interference patterns which result in intensity variations of received acoustic signals. This presentation provides an overview of the R/V Sharp's transmissions, and the corresponding intensity fluctuations of received signals at the array. These fluctuations are compared to internal wave events that were recorded at both the transmission and reception locations. These internal wave events were also imaged by the R/V Sharp's radar and satellite-based radar. Following the work of Fredricks *et al.* [J. Acoust. Soc. Am. **117**, 1038 (2005)], statistical distributions are fit to the calculated intensities for different transmission directions, distances, and times. These distributions are compared to modeled data with and without the internal wave field. [Work sponsored by the Office of Naval Research.]

2:50

2pAO7. Time-frequency pattern of the sound intensity fluctuations of midfrequency signals in presence of internal waves in Shallow Water 06 experiment. Boris Katsnelson, Valery Grigorev, and James Lynch (Woods Hole Oceanograph. Inst., 98 Water St., Woods Hole, MA 02543, jlynch @whoi.edu)

In given paper fluctuations of intensity of sound signals, radiated by the midfrequency source (R/V Knorr) during ~5 h were studied. Broadband signals (2–8 KHz) were received by four single hydrophones fixed at the bottom (SHRUs) placed at different distances from the source (from 4 to 12 km). These four acoustic tracks have different directions relative direction of propagation of the train of intensive internal waves (the corresponding angles in horizontal plane are from ~5 to ~15 deg). Time-frequency diagrams were constructed using frequency filtering of the spectrum of broadband signals with sliding narrow window. Temporal dependences of intensities of received signals within frequency window were constructed for period 14:00 until 19:00 GMT for experiment carried out on 13 August. During this time train of internal solitons was registered, propagating toward the coast. Mentioned time-frequency diagrams demonstrate specific features of influence of internal waves on the temporal variations of the sound intensities at four SHRUs. More exactly variations of predominating frequency in spectra correspond to variations of positions of the solitons at the acoustic tracks. Results of experimental data are compared with theoretical estimations. [Work was supported by RFBR and CRDF.]

3:05

2pAO8. Acoustic ducting, refracting, and shadowing by curved nonlinear internal waves in shallow water. James F. Lynch, Ying-Tsong Lin, Timothy F. Duda, Arthur E. Newhall, and Glen Gawarkiewicz (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543, jlynch@whoi.edu)

Nonlinear internal waves are common on continental shelves and have been shown to have strong effects on acoustic propagation and scattering. Transmissions from previous field work performed in an along-shelf geometry show very strong ducting of low-frequency (50–500 Hz) sound between nonlinear waves. This strong ducting effect is acoustically important. Most all of the acoustic propagation studies of internal waves have used straight line internal wave fronts for individual waves or packets. In our work, we implement a theoretical analysis using the Weinberg-Burridge horizontal rays and vertical modes formalism to study the acoustic ducting, refracting and shadowing of low-frequency sound due to curved internal waves, which seem to dominate in shelf break regions. A three-dimensional parabolic-approximation sound propagation model is also used with consideration of realistic environmental conditions, and the modeling results provide a clear depiction of the underlying physical processes. The oceanographic origin of the internal wave curvature is discussed, as this affects the predictability of the sound speed field. [This work is sponsored by the Office of Naval Research.]

Session 2pBB

Biomedical Ultrasound/Bioresponse to Vibration: Biomedical Applications of Standing Waves

Armen Sarvazyan, Cochair
ARTANN Laboratories, 1459 Lower Ferry Rd., Trenton, NJ 08618

Martin Wiklund, Cochair
Dept. of Applied Phyiscs, Royal Inst. of Tech., SE-10691, Stockholm, Sweden

Chair's Introduction—1:15

Invited Papers

1:20

2pBB1. Ultrasonic manipulation in microfluidic chips for accurate bioparticle handling. M. Wiklund, O. Manneberg, J. Svennebring, B. Vanherberghen, B. Onfelt, and H. M. Hertz (Dept. of Appl. Phys., Royal Inst. of Technol., SE-10691 Stockholm, Sweden)

Micro-manipulation of cells or other bioparticles has important applications in biological and biomedical research. Dielectrophoresis and optical tweezers are the classical tools in this field. Ultrasonic standing wave (USW) technology is presently emerging as powerful alternative, especially in microfluidic chips and other miniaturized systems. In this contribution we will review our present activities in USW-based particle manipulation in microfluidic chips. Recent experiments and simulations allow us to tailor the spatial distribution of the USW force field by multi-frequency actuation and appropriate microchannel/transducer resonance design. We introduce the concept of ultrasonic micro-cages for single-cell or single-particle 3-D ultrasonic manipulation. It is shown that these manipulation tools can be combined with high-resolution optical microscopy, thereby allowing state-of-the-art characterization of individual cells. Frequency-modulation is shown to stabilize the manipulation performance as well as allowing flow-free transport of particles and cells. Environmental control inside the channels has been achieved and proliferation and viability studies are promising. Finally, for bio-analysis, we demonstrate potential for femtomolar bio-analytical sensitivity in bead-based assays using USW enrichment.

1:40

2pBB2. Acoustofluidics: Theory and simulation of streaming and radiation forces at ultrasound resonances in microfluidic devices. Henrik Bruus (Dept. of Micro- and Nanotechnology, Tech. Univ. of Denmark, DTU Nanotech, Bldg. 345 East, DK-2800 Kongens Lyngby, Denmark, Henrik.Bruus@nanotech.dtu.dk)

During the past few years, there has been an increasing interest in applying ultrasound waves to manipulate biological particles and liquids in microfluidic devices. To obtain optimized designs and functionalities of the acoustofluidic devices, more detailed theoretical studies and numerical simulations are called for. The basic second-order perturbation theory is presented for acoustic fields applied at ultrasound frequencies in silicon/glass systems containing water-filled microfluidic channels and chambers. For various specific device geometries, the resonance frequencies and corresponding modes of the acoustic fields are calculated numerically to first order. At these frequencies, the largest possible acoustic powers are obtained in the microfluidic system. The first order fields are then used as source terms in the equations for the time-averaged second order pressure and velocity fields, which are directly related to the acoustic radiation force on single particles and to the acoustic streaming of the liquid. For the radiation pressure effects, there is good agreement between theory and simulation, while the numeric results for the acoustic streaming effects are more problematic. Possible improvements in the latter case are discussed.

2:00

2pBB3. Transient processes in acoustic resonators. Armen Sarvazyan (Artann Labs., 1753 Linvale-Harbourton Rd., Lambertville, NJ 08530, armen@artannlabs.com) and Lev Ostrovsky (Zel Technologies/Univ. of Colorado, Boulder, CO 80305)

The use of ultrasound in resonators for manipulating particles in various biomedical applications is a relatively well studied topic. These studies were mostly concerned with steady-state processes. However, in certain important applications, such as stirring and mixing fluids in microfluidics and biosensors, the time scale of a process plays a crucial role. In this presentation we consider some of these applications. In particular, the radiation force-induced motion of microparticles and microbubbles in a swept-frequency ultrasonic resonator is considered. The particles are forced to move due to switching the resonance modes in a resonator cell, thus providing effective stirring of the fluid. Another field of medical application of standing waves, where transient processes are of crucial importance, is ultrasonic treatment of tissues. In this new field of ultrasonic therapy, standing waves are used for producing lesions in the tissue with simultaneous monitoring the formation of the lesion by automatic controlling parameters of the standing wave field. The spatial and temporal distribution of ultrasound energy in the resonator containing treated tissue portion is theoretically analyzed. Dynamics of temperature changes in the tissue, which is important for optimizing tissue treatment regimes, is evaluated. Experimental data illustrating theoretical results are presented.

2:20

2pBB4. Current applications of ultrasonic resonators to biomolecular studies. Tigran Chalikian (Faculty of Pharmacy, Univ. of Toronto, 144 College St., Toronto, ON M5S 3M2, Canada, chalikan@phm.utoronto.ca)

The most versatile family of instruments providing the highest precision of acoustic measurements in the lowest sample volumes utilizes the method of fixed-path interferometer. The speed of sound in a medium is a simple function of its density and compressibility. Both parameters are determined by and, therefore, reflect the entire spectrum of intra- and intermolecular interactions within the system. In addition, compressibility probes the pressure response of these interactions. Current resonator-based instruments provide reliable measurements in diluted solutions of biological compounds with solute concentrations on the order of 1 mg/mL and solution volumes of 1 mL. They require less than 1 mg of sample, a range acceptable for the majority of biochemical and biophysical investigations. Ultrasonic velocimetric measurements have been employed to study a range of biologically relevant reactions, including the folding /unfolding transitions of proteins and helix-to-coil transitions of nucleic acids, as well as drug-DNA, ligand-protein, protein-protein, and protein-DNA association events. The limiting factor in such investigations is the need to rationalize the measured properties (volume, compressibility, and their derivatives) in terms of microscopic events, such as solute-solvent interactions. This presentation reviews biomolecular applications of acoustic measurements and discusses the development of approaches for microscopic interpretation of macroscopic volumetric observables.

2:40

2pBB5. Cylindrical standing wave resonator for liquid food quality control. Aba Priev (Hebrew Univ.-Hadassah Med. School, Jerusalem 91120; NDT Instruments Ltd., 56 Anilevich St., Jerusalem, Israel, abbap@ekmd.huji.ac.il) and Armen Sarvazyan (ARTANN Labs., Inc., Lambertville, NJ 08530)

In this paper, an innovative technology based on the use of ultrasonic cylindrical standing waves for continuous monitoring of quality of various liquid food products, such as milk, juices, beer, wine, and drinking water is described. A proprietary unique feature of the developed ultrasonic analyzer is that it employs a combined mode of operation using both high-intensity and low-frequency (10 W/cm^2, 1 MHz) waves for separation and concentration of the high-molecular-weight particles (fat globules or cells) and low-intensity and high-frequency (0.5 W/cm^2, 10 MHz) waves for compositional analysis. High accuracy for ultrasound velocity measurements (up to 0.001%) and ultrasound attenuation (of about 1%) and rapid testing time (2–20 s) have been achieved. Comparative analyses of the ultrasonic method with standard reference techniques have produced linear calibration curves for major components with correlation coefficients higher than 0.95. It is thus possible to monitor total protein and fat content, and somatic cell count in raw milk in cowsheds, or salinity, turbidity, specific gravity, and particles (bacteria) in drinking water directly. Advantages of the proposed technology include the reagent-free nature, no need for sample pretreatment, ease-of-use, and low cost.

3:00—3:30 Break

Contributed Papers

3:30

2pBB6. Movement of liquids and solid particles on multiwall acoustic plates. Jeremy J. Hawkes (School of Chem. Eng. and Analytical Sci., The Univ. of Manchester, MIB 131 Princess St., Manchester, M1 7DN, UK), Rito Mijarez-Castro (Instituto de Investigaciones Elctricas, Morelos, Mexico, rmijarez@iie.org.mx), and Peter R. Fielden (The Univ. of Manchester, Manchester, M1 7DN, UK)

Well-defined vibration modes are used to form acoustic wells on flat plates. Within these dynamic traps, acoustic streaming continually stirs the contents. Traps can be used to hold liquids in separate heaps on a surface or to hold particles in levitated arrays. A protocol for creating trap patterns will be described, where FE modelling is used to select modes and then maneuver the modes to the desired frequency.

3:45

2pBB7. Nonlinear ultrasonic standing waves in bubbly liquid: A numerical study. Vanhille Christian (Escet, Univ. Rey Juan Carlos, Tulipán s/n, 28933 Móstoles, Madrid, Spain, christian.vanhille@urjc.es) and Campos-Pozuelo Cleofé (Inst. de Acústica, CSIC, 28006 Madrid, Spain)

This study deals with the behavior of nonlinear ultrasonic standing waves in a bubbly liquid. A water air mixture is considered. The SNOW-BL numerical code gives the solution of a differential system coupling the nonlinear bubble equation with the linear wave equation. Several kinds of high power standing waves which may be useful in biomedical applications are considered here: pulses propagating in water with air bubbles clouds or layers, harmonic waves in a bubbly resonator, non-resonant standing waves in a fluid with bubbles. Conclusions about acoustic pressure distributions, waveforms, and frequency content, as well as mean and rms pressures, are given. Particular acoustical effects induced by the dynamics of the air-bubbles on the pressure standing wave are commented. Comparisons with results in homogeneous fluids are also presented. [Work supported by DPI2008-01429.]

4:00

2pBB8. Use of resonator method for body fluids composition assessment. Viktor Klemin (BIOM, 3 Veterinarnaya Str., Nizhny Novgorod, 603098, Russia, klemina@rf.unn.ru)

Ultrasonic resonator method of body fluids composition evaluation is described. The experimental data on assessment of whole blood, blood serum, gastric juice, and saliva are presented. The developed method is based on the assumption of additivity of the contributions of major components of fluids in their acoustic characteristics. The acoustic characteristics measured by the ultrasonic resonator method include the speed and absorption of ultrasound and their frequency and temperature dependences. The device for evaluation of body fluid composition comprises differential ultrasonic resonator cells with the volume of 80 mkl each, operating at about 8 MHz, mounted in a micro-thermostat, and a compact electronic unit. The method allows for simultaneous determination of total protein, $Na+$ and $K+$, in saliva and gastric juice, lipid components (cholesterol total, cholesterol of high and low density, and triglycerides), the total protein and protein fractions in blood serum within a very short period of time, on the order of minutes, in contrast to conventional time consuming tests. Other advantages of the proposed technology include the reagent-free nature, no need for sample pretreatment, ease-of-use, and low cost. The accuracy of ultrasonic method for body fluid composition assessment is comparable to that of conventional time-consuming and expensive assays.

4:15

2pBB9. Ultrasound standing wave fields control the spatial distribution of cells and protein in three-dimensional engineered tissue. Kelley A. Garvin (Dept. of Biomedical Eng., 205 Goergen Hall, Univ. of Rochester, Rochester, NY 14627, garvin@bme.rochester.edu), Denise C. Hocking, and Diane Dalecki (Univ. of Rochester, Rochester, NY 14627)

The application of ultrasound standing wave fields (USWFs) to cell suspensions results in the radiation force-mediated movement of cells to areas of the field separated by half-wavelength intervals. In this study, USWF-induced cellular arrangements were maintained after removal of the sound field by polymerization of a collagen type-I solution around the organized cell bands. Using a water tank setup, cell suspensions of varying concentrations were exposed to a 1-MHz, continuous wave USWF for 15 min, during which time collagen polymerization occurred. An acoustic pressure amplitude of 0.2 MPa was used to achieve the USWF-induced cell organization. Compared to sham samples with a random cell distribution, cell viability was not adversely affected by USWF exposure. The organization of cells into a banded pattern within the collagen gels resulted in a significant two-fold increase in cell-mediated gel contraction, suggesting that USWF-induced cell organization leads to differential extracellular matrix remodeling. Protein organization within the tissue constructs was further controlled by USWF-mediated collocalization of soluble fibronectin to cell bands. These technologies have applications to the fabrication of tissue analogs with desired tissue characteristics for the repair or replacement of diseased or injured tissues and organs in the field of tissue engineering.

4:30

2pBB10. Two-dimensional numerical simulations for the time-averaged acoustic forces acting on a rigid particle of arbitrary shape in a standing wave. Jingtao Wang (Ctr. of Mech., Dept. of Mech. and Process Eng., ETH Zurich, 8092 Zurich, Switzerland, wang@imes.mavt.ethz.ch) and Jurg Dual (ETH Zurich, 8092 Zurich, Switzerland)

The time-averaged acoustic force can be applied to many practical fields such as contactless particle manipulation in biomedicine. It is necessary to accurately predict the mean forces on suspended obstacles to design ultrasonic particle manipulators. Although there have been many analytical solutions on this topic, it is difficult to determine the acoustic forces on obstacles under more complex system conditions such as proximity to the chamber wall, complex viscous function, acoustic streaming, and complicated particle shapes. Therefore, the numerical modeling may become a powerful tool. In this paper, the time-averaged forces, which act on rigid 2-D particles with different shapes in ideal and viscous fluids exerted by a standing sound wave field, are computed by solving the Navier-Stokes equations directly using the finite volume method (FVM) technique. The cylinder results agree well with Haydock's theoretical prediction and his lattice Boltzmann simulations. Then, the force and torque acting on a needle shaped particle in a standing wave are calculated and discussed in detail. Furthermore, the viscous effects of the host medium are also investigated. Our program with the FVM algorithm proves to be quite suitable for calculating the acoustic forces in standing waves.

4:45

2pBB11. Time reversal of monochromatic signals in elastic media. Brian E. Anderson, Robert A. Guyer, Timothy J. Ulrich, and Paul A. Johnson (Geophys. Group EES-17, Los Alamos Natl. Lab., Los Alamos, NM 87545)

A set of experiments has been conducted to show that time reversal of steady state monochromatic signals can produce spatial focusing in a reverberant elastic cavity when multiple channels are used. The transient portion of the received signals is not used. A single channel does not produce spatial focusing as it only drives the system according to its modal distribution. The amplitude of the energy at the focal location increases as the square of the number of channels used, while the amplitude elsewhere in the medium increases proportionally with the number of channels used. This work has importance in the field of medical ultrasound where the use of a long duration monochromatic excitation may be used for lithotripsy or other ultrasonic therapy. [This work is supported by Institutional Support (LDRD) at Los Alamos National Laboratory.]

TUESDAY AFTERNOON, 19 MAY 2009 COUNCIL SUITE, 1:00 TO 3:30 P.M.

Session 2pEAa

Engineering Acoustics: Emerging Transduction Devices

Fletcher Blackmon, Chair
Naval Sea Systems Command Division Newport, Newport, RI 02841

Contributed Papers

1:00

2pEAa1. The piezoelectric thickness mode hydrophone using tonpilz structure for deep-sea application. Min Sung, Haksue Lee, and Wonkyu Moon (Dept. of Mech. Eng., Pohang Univ. of Sci. and Tech., PIRO, Rm. 416, san31 hyja-dong, nam-gu, Pohang-si, kyungsangbuk-do, Korea)

A piezoelectric thickness mode hydrophone for deep-sea application using tonpilz structure was designed, and its macroprototype was made as a preliminary for the micromachined version. Several types of micromachined thin-diaphragm sensors for hydrophone application have been reported. They have two issues in deep-sea applications. One is a sensitivity problem. To enhance the sensitivity, a cone structure for sound pressure amplification was placed between the clamped diaphragm and the piezoelectric square plate. To evaluate the effect of the cone structure on sensitivity, which was motivated by tonpilz structure, an air-backed prototype was made using aluminum and piezoelectric ceramic square plate of 700 μm length and 400 μm thickness. A lumped parameter model was used in the design procedure. The measured sensitivity of the prototype was −249.60 dB ref. One volt per micropascal without a preamplifier, and the measured resonant frequency was 30 kHz, which agreed with the prediction. The other is an endurance problem due to high hydrostatic pressure. To develop the endurance for the high hydrostatic pressure in deep-sea, PZT thickness mode and oil-backing on the diaphragm backside were proposed. The comparison of oil-backing and air-backing is to be presented as an evaluation of the endurance development. [Research supported by MRCnd.]

1:15

2pEAa2. Rapid identification of candidate materials for tonpilz head-mass design. Scott P. Porter, Douglas C. Markley, David J. Van Tol, and Richard J. Meyer, Jr (Appl. Res. Lab., The Pennsylvania State Univ., P. O. Box 30, State College, PA 16804, scott.porter@psu.edu)

Optimum performance of a tonpilz transducer head-mass section is often achieved when the mass is minimized and the stiffness is maximized. High stiffness keeps head-mass modes away from the band of interest with minimal material; low mass maximizes the motor section aspect ratio and re-

duces the stored energy in the head-mass, raising the electromechanical coupling coefficient. Therefore, the material selected for the head-mass construction plays an integral role in transducer performance. In the practical design of tonpilz radiating heads, a tool which can rapidly identify the materials best suited for a given design is desirable. The figure of merit, $E\rho^3$, has traditionally been used. This figure of merit, however, is based on thin-plate approximations. Most actual tonpilz head-masses require thick-plate theory for accurate analysis. In this paper, the authors use a numerical routine to rapidly assess the thick-plate behavior of various materials and rank their suitability for a given head-mass geometry. These results are corroborated with finite element modeling. Finally, the thick-plate ranking scheme is compared to the traditional figure of merit. The results of this comparison challenge established assumptions about the best materials for head-mass design and identify ceramic materials as strong candidates for future designs.

1:30

2pEAa3. Investigation of single crystals for U.S. Navy standard electro-acoustic transducer applications. Jeffrey A. Szelag and Thomas R. Howarth (Naval Sea Systems Command Div. Newport, Newport, RI)

The Underwater Sound Reference Division (USRD), Naval Sea Systems Division in Newport, RI, is the U.S. Navy's primary source for the provision of standard underwater electro-acoustic reference transducers. USRD has over 40 years of traceable and repeatable historic data for each transducer that is currently available for underwater calibrations and experiments. As such, USRD's primary focus is on the continuous development of new reference transducers incorporating the latest technologies available. Requirements for these standards include repeatability over extended time, traceability, and ease of maintenance. This presentation covers an on-going USRD investigation to feature single crystal PMN-PT as the active substrate in the current F41 program transducer. In addition, lossy rubber substrates are being examined to replace the legacy corprene passive materials in order to extend the operating depths of the transducer. Our presentation will begin with an overview of the fabrication techniques followed by experimental underwater data comparing the new active and passive materials with the traditional F41.

1:45

2pEAa4. A generalized differential effective medium model of piezoelectret foams. Michael R. Haberman (Appl. Res. Labs., The Univ. of Texas at Austin, 10000 Burnet Rd., Austin, TX 78758)

Piezoelectrets are closed-cell foams consisting of a continuous polymer matrix containing preferentially oriented and electrically polarized ellipsoidal voids. These materials display low density, low characteristic impedance, thicknesses on the order of 100 μm, and potential to act as conformal transducers on complex geometries. They therefore provide an intriguing alternative to piezoelectric ceramics and piezopolymers for applications such as air-coupled ultrasonics or underwater receive elements. Unfortunately, some fundamental attributes of piezoelectret behavior, such as their nonmonotonic change in receive sensitivity with increasing ambient pressure, remain unexplained by physical modeling or experiment. To explain this and other behavior, this work develops a fully coupled multiscale microelectromechanical model which captures the void-scale physics and predicts macroscopically observed piezoelectric properties. The model employs volume averaged Green's function solutions of the three-dimensional stress equilibrium equations and Gauss's Law. The resulting model approximates the effective stiffness, dielectric permittivity, and piezoelectric coupling tensors as a function of the constituent material properties, void fraction, shape, and orientation, as well as the deposited charge density. Strain localization tensors of the model are then employed to approximate piezoelectret dependence and external pressure. Model results show good agreement with experimental results and data from the literature.

2:00

2pEAa5. A micromachined piezoelectric flexural mode hydrophone for audible frequency application. Haksue Lee, Sungjoon Choi, and Wonkyu Moon (Dept. of Mech. Eng., Postech, South Korea, wkmoon@postech.ac.kr)

A micromachined hydrophone is designed for audible frequency application (20 Hz–20 kHz). The basic structure is a flexural unimorph consisting of a clamped silicon diaphragm and a piezoelectric film. Two design approaches can enhance its sensitivity. One is a unimorph with unequal radii between the piezoelectric film and the diaphragm, which increases electromechanical coupling. The other is an air-backed chamber, which reduces mechanical impedance in the flexural vibration and, consequently, enhances sensitivity. To predict the effects of these approaches on sensitivity, an equivalent transduction model is composed and used for calculations. In the case of oil backing, compared to air backing, a reduction in sensitivity of about 40 dB is expected. The dimensions of the micromachined diaphragm (piezoelectric film) are 630 μm (400 μm) radius and 15 μm (3 μm) thick. A fabricated microsensor is constructed as a hydrophone by using a rubber housing filled with castor oil. The measured sensitivity is −228.3 dB at 1V /μPa without a preamplifier. A flat frequency response within ±1 dB deviations were measured in the 100 Hz–5 kHz band. Based on the simulation, it can resist the hydrostatic pressure of 100 m depth, without pressure balancing, while decreasing in sensitivity. The results of a larger model having a higher sensitivity of −207.6 dB will also be presented. [Research supported by MRCND and R0A-2007-000-20042-0.]

2:15

2pEAa6. Advantages of piezoelectric microelectromechanical systems (MEMS) microphones. Robert Littrell and Karl Grosh (2250 GG Brown Bldg., 2350 Hayward St., Ann Arbor, MI 48109)

Microphones fabricated using microelectromechanical systems (MEMS) technology are one of the fastest growing applications of MEMS. Capacitive sensing has been the dominant form of transduction in both traditional and recently commercialized MEMS microphones. Models and experiments, however, indicate that the thin layers and fine spatial resolution made possible by MEMS technology lend themselves more appropriately to piezoelectric microphones. Although piezoelectric MEMS microphones have typically been shown to have a relatively high noise floor, this limitation can be overcome with appropriate design and high quality piezoelectric material. Models indicate that piezoelectric MEMS microphones can achieve a comparable noise floor to capacitive MEMS microphones of similar size and bandwidth while achieving 1000 times greater dynamic range. A piezoelectric MEMS microphone utilizing aluminum nitride (AlN) will be presented. Previously experienced film quality issues have been addressed. This microphone is designed to have a noise floor below 40 dB (A), a dynamic range greater than 160 dB SPL, a 10 kHz bandwidth, and a sensing area of less than 1 mm^2.

2:30

2pEAa7. Infrasonic microphone. Timothy Marston and Thomas B. Gabrielson (Grad. Prog. in Acoust., Penn State Univ., PO Box 30, State College, PA 16804, tmm357@psu.edu)

The high cost of many infrasonic transducers can be a setback for universities interested in performing research in the field of infrasound. Reasonably effective infrasonic transducers can be constructed in the laboratory, however, and for a fraction of the cost. A simple, inexpensive infrasonic microphone that operates on the principle of carrier demodulation has been constructed. Noise analysis of the microphone demonstrates its potential usefulness for cost-effective field research and deployment in temporary infrasonic arrays. [Research funded by the Penn State Applied Research Laboratory Educational and Foundational Fund.]

2:45

2pEAa8. Mechanical characterization of capacitor microphones for analysis and study. YuFan Chuang (Master Program of Electroacoustics, Feng Chia Univ., 100 Wenhwa Rd., Seatwen, Taichung, Taiwan 40724, ROC, yufan420@gmail.com), S. J. Pawar, and Jin H Huang (Feng Chia Univ., Taichung, Taiwan 40724, R.O.C.)

An attempt is made to investigate the mechanical properties of the diaphragm of capacitor microphone using MTS Tytron 250 micro-force testing machine. The compliance of the diaphragm is calculated from the force-displacement diagram. The simulation of the microphone is carried out using equivalent circuit method. The mechanical parameters like mass

(measured), compliance (measured and calculated), and resistance (assumed from literature) are being used to formulate mechanical domain of microphone in equivalent circuit model. The sensitivity curve of microphone is being plotted. Simultaneously, the experimental measurement has been carried by PULSE Electroacoustics with the software Sound Check in an anechoic chamber. The simulation and experimental results for sensitivity are found to follow each other within a wide frequency range. The characteristic investigation is carried out using validated equivalent circuit model for parametric study of microphone. [Work supported by Contract Nos. NSC 95-2221-E-035-013-MY3 and 95-2221-E-035-087. The authors would like to thank Leo Liao and Jack Wei of Merry Electronics Co. Taiwan and Listen Inc. for supplying Sound Check measurement.]

3:00

2pEAa9. Head related transfer functions for KEMAR. Gunnar Rasmussen (G.R.A.S. Sound & Vib., Skovlytoften 33, 2840 Holte, Denmark, gr@gras.dk)

The head related transfer functions (HRTFs) for KEMAR for a number of ear pinnas of different sizes and hardnesses have been measured. Right and left ear differences on KEMAR as well as change in HRTFs over the last 40 years will be reported along with data for the ITU - Type 3 and the IEC 60959 standardized pinna. The influencing factors on the uncertainty budget will be discussed.

3:15

2pEAa10. A transfer matrix method for estimating the dispersion and attenuation of plane waves in the standing wave tube. Kang Hou and J. Stuart Bolton (Herrick Lab, Purdue Univ., West Lafayette, IN)

In this paper, an iterative method based on the transfer matrix approach is described for evaluating the sound speed and attenuation constant of air in a standing wave tube. The procedure is based on transfer matrix, the four-microphone method for measuring the transmission loss of acoustical materials. In the latter procedure, a sample is placed between two pairs of microphones, where the sample is exposed to an incident plane wave field. In the present case, the air inside the standing wave tube is treated as the sampled material. Since the knowledge of the dispersion characteristics of plane waves within the tube is required to perform this measurement, an iterative method must be applied at the postprocessing step to estimate the complex wave numbers in the sample section. The result from the experiments showed that the air temperature within the tube has a significant impact on the results. The measured results were found to be in good agreement with the Temkin formula for attenuation in tubes. In addition, the approach may be extended to the accurate measurement of viscosity and sound velocity of liquids and gases.

TUESDAY AFTERNOON, 19 MAY 2009 COUNCIL SUITE, 3:30 TO 4:40 P.M.

Session 2pEAb

Engineering Acoustics: Piezoelectric Energy Harvesting

Thomas R. Howarth, Chair
Naval Sea Systems Command Division Newport, Newport, RI 02841

Chair's Introduction—3:30

Invited Papers

3:35

2pEAb1. Bridging the gap between macro- and microdevices for manufacture of piezoelectric energy harvesters. Arthur L. Chait (EoPlex Technologies, Inc., 3698-A Haven Ave., Redwood City, CA 94063, achait@eoplex.com)

In this presentation, materials engineer and entrepreneur, Arthur L. Chait will show why there is a gap at the miniature scale when building complex devices with multiple materials. He will explain why few technologies can handle several materials, with complex geometries, in small parts. Miniature piezoelectric (PZ) energy harvesters are an example of this challenge. Mr. Chait will describe his piezoelectric energy harvester, designed for tire pressure sensor systems. This tiny device uses PZ materials to convert some of the vibration of the tire into electricity. The electric current is stored in a capacitor and used to power the system. Audiences will learn why it is difficult and costly to build a small, rugged system to power a tire sensor. The anatomy of this device includes at many diverse materials and five basic design elements. These are bonded into a strong package about the size of a contact lens. Building this energy harvester would be almost impossible with conventional manufacturing techniques. Mr. Chait will share a relatively new solution; i.e., a print deposition technique that produces thousands of parts simultaneously from many different materials. Active elements like circuits, catalyst beds, mixing chambers, capacitors, and piezoelectric actuators are produced all in one step.

4:00

2pEAb2. Acoustic energy harvesting using an electromechancal Helmholtz resonator. Fei Liu (Dept. of Mech. and Aerosp. Eng., Univ. of Florida, Gainesville, FL 32611-6250), Alex Phipps , Stephen Horowitz, Louis Cattafesta , Toshikazu Nishida , and Mark Sheplak (Univ. of Florida, Gainesville, FL 32611-6250)

This talk presents the development of an acoustic energy harvester using an electromechanical Helmholtz resonator (EMHR). The EMHR consists of an orifice, cavity, and a piezoelectric diaphragm. When the acoustic wave is incident on the EMHR, a portion of acoustic energy is converted to electrical energy via piezoelectric transduction in the diaphragm of the EMHR. Moreover, the diaphragm is coupled with energy reclamation circuitry to increase the efficiency of the energy conversion. Two power converter topologies are

adopted to demonstrate the feasibility of acoustic energy reclamation using an EMHR. The first is comprised of only a rectifier, and the second uses a rectifier connected to a flyback converter to improve load matching. Experimental results indicate that approximately 30 mW of output power is harvested for an incident sound pressure level of 160 dB with a flyback converter. Such power level is sufficient to power a variety of low power electronic devices.

Contributed Paper

4:25

2pEAb3. Theoretical study of thermoacoustic power conversion with a piezoelectric transducer. Carl Jensen and Richard Raspet (Natl. Ctr. for Physical Acoust., Univ. of MS, Univ., MS 38677)

A heat-driven thermoacoustic engine prototype has been designed with a piezoelectric transducer as an electroacoustic transformer. The design under consideration uses a bending, unimorph transducer placed between two identical thermoacoustic engines operating out of phase such that the transducer is driven in a push-pull fashion. The thermoacoustic engine is modeled using a one-dimensional low amplitude simulation. The piezoelectric unimorph is modeled by an electroacoustic equivalent circuit aimed at capturing the first order resonant behavior of the transducer; loss mechanisms from hysteretic behavior of the dielectric or elastic properties are not considered by the model. Issues in optimizing the geometry of the model will be presented as well as comparisons to other forms of waste heat power harvesting such as thermoelectric.

TUESDAY AFTERNOON, 19 MAY 2009 STUDIO SUITE, 1:00 TO 4:00 P.M.

Session 2pMUa

Musical Acoustics: Wind Instruments II

R. Dean Ayers, Chair
624 Valley View Dr., Medford, OR 97504-6356

Invited Papers

1:00

2pMUa1. Evaluation of two control parameters of trumpet players as function of sound features. Rene Causse (Ircam, 1 place Igor Stravinsky, F-75004 Paris, France) and Vincent Freour (McGill Univ., Montréal, PQ Canada.)

Experiments were performed by five trumpet players with the same trumpet and mouthpiece. For each subject, two control parameters are measured: mouth-pressure and force applied by the lips on the mouthpiece. The sound is also recorded. The players are asked to play sustained notes at different pitch and sound levels. We calculate fundamental frequency, sound pressure level, and some spectral features: noisiness, spectral roll-off, etc. Although little mouth-pressure variability as function of pitch and dynamic appears between subjects, clear differences were obtained in terms of the force applied on the mouthpiece. Therefore, we suggest to focus on the link between sound features and the ratio of these two control parameters measured, sort of "gesture estimator." The representation of the ratio as function of the sound features shows a good classification of musicians. Moreover, in some specific cases such as noisiness, the relation seems independent from the players. These observations suggest that these two control parameters could be efficient to characterize players' embouchure and to access some sound characteristics without knowledge of the mechanical characteristics of the lips. These results will be described and discussed, and direction for future works suggested. [Consonnes project, supported by the French Research Agency.]

1:30

2pMUa2. Does a brass-instrument's timbre depend on the alloy from which it is made? Robert W. Pyle, Jr. (11 Holworthy Pl. Cambridge, MA 02138, rpyle@post.harvard.edu)

Most brass players with a few years experience would give an affirmative answer to the question posed in the title. In the past, audible differences have proven elusive when pursued through spectral analysis of sustained tones, perhaps because the right question was not asked. This talk will examine several attributes of trombone spectra. The trombone used has a "modular construction" so that different bells can easily be installed on an otherwise unchanged instrument. A clear difference, consistent with the player's and listener's perception of the tone quality, emerges between yellow brass (70% Cu, 30% Zn) and red brass (90% Cu, 10% Zn). Transient behavior, particularly during the attack, is at least as important to the performer as sustained sounds. Here, too, differences are attributed to the alloy. Yellow brass is characterized by "crisp" attacks, red brass by "rounded" attacks. This talk will also analyze attack transients, even though attacks that are perceived as similar appear quite variable when examined in detail.

2:00

2pMUa3. Comparison of the mechanics of the brass players' lips during slurred note transients. John P. Chick and Shona M. Logie (SEE, Univ. of Edinburgh, Kings Bldgs., Edinburgh EH9 JL, UK, john.chick@ed.ac.uk)

When sounding a note on a brass instrument, a strong coupling is established between the vibrating lips of the player and the air column resonance. In moving between notes as smoothly as possible, or "slurring," there is a transition from one strongly coupled, steady-state, regime of the lip/air column to a different steady-state coupling of the lip/air column. To the listener, there are sometimes subtle differences in the sound of the transition between notes depending on how this transition is achieved: by lip slur, valve slurs, and, in the case of the trombone, slide slurs. This paper uses data collected from high speed video capture of the players' lips, synchronized with microphones in the mouthpiece and the bell of the instrument, to investigate and compare the motion of the brass players' lips for different types of slurred internote transients.

2:30

2pMUa4. Acoustical history of the tuba. Arnold Myers (Univ. of Edinburgh, Reid Concert Hall, Bristo Square, Edinburgh EH8 9AG, U.K.)

Low-pitched valved brass wind instruments are often described generically as tubas. In fact, the term "tuba" covers a broad family of instruments, with sounding lengths ranging from the 8 ft C of the classic French orchestral tuba to the 18 ft Bb of the brass band and orchestral contrabass. The first instrument designated as a tuba was introduced in 1835, and since then a variety of models differing in bore profile as well as in nominal pitch have been used. This paper explores the the historical development of the tuba family from an acoustical perspective, presenting and discussing measurements of bore profile and input impedance for a number of representative instruments from different periods and musical traditions.

3:00—3:15 Break

Contributed Papers

3:15

2pMUa5. The double bell descant euphonium. Frederick J. Young (Div. of Com. and Arts, Univ. of Pittsburgh, 800 Minard Run Rd., Bradford, PA)

A euphonium is presented that has better intonation and is easier to play in the extreme registers. It is a complete double horn having five double valves in contrast to the incomplete normal double French horn having only three double valves. The valves descend 2, 1, 3, 4, and 5 semitones from the open tones. The large bell is for the B♭ standard part of the instrument. A switch valve activates the smaller alto bell in the key of high E ♮ and sends the air through the five shorter valve slides and out the small bell. The fundamental frequency of the E side is 82.4 Hz or six semitones above the fundamental of the B♭ euphonium. The key of E rather than F or E♭ enables all notes in the complete range (about 29–988 Hz) to be played with a single tunable valve slide. Either of the other keys would introduce a need for one additional valve or compromise the intonation.

3:30

2pMUa6. Optimization of valve tube lengths for brass instruments. Frederick Young (Div. of Com. and Arts, Univ. of Pittsburgh, 800 Minard Run Rd., Bradford, PA 16701, youngfj@youngbros.com)

The intonation deviations from the well-tempered scale are presented for valve brass wind instruments. Compensating and full double instruments are considered. Systems including descending valves and less popular ascending valves are studied. The lengths of the valve tubes are optimized by minimizing the overall root mean squared (rms) intonation error using the method established in 1967 [R. Young, J. Acoust. Soc. Am. **42**, 224–235 (1967)]. Open note detuning is considered in each case and improves the overall intonation. The worst three valve instrument without open note detuning has an rms deviation 24 cents The best of the three and four valve instruments is a three valve compensating euphonium tuned 3 cents sharp on the open notes and flat on valves 2, 1, and 3 by 3, 3, 10 cents, respectively. It exhibits only 2 cents rms deviation on single valves and less than 3 cents on combinations of valves. The comparisons also include several new three, four, and five valve systems. Included is a double horn with only two double valves and a switch valve.

3:45

2pMUa7. The input impedance of an alphorn including an Alexander mouthpiece. Frederick J. Young (Div. of Com. and Arts, Univ. of Pittsburgh, 800 Minard Run Rd., Bradford, PA 16701, youngfj@youngbros.com)

The alphorns built and sold by the late Joseph Littleton of Hammondsport, NY are used in this study. His dimensions of bore as a function of position are used in solution of the Webster equation for impedance solved numerically earlier [F. J. Young, Acustica, **10**, 91–97 (1960)]. The musical notation for the frequencies of the impedance peaks are given below. Here the subscripts refer to the harmonic number. The fundamental, rarely used, is 13 cents flat from A at 55 Hz. The intonations errors for. C_1, C_2, G_3, C_4, E_5, G_6, B^{b}_{7}, C_8, D_9, E_{10}, F_{11}, G_{12}, A_{13}, B^{b}_{14}, B_{15}, C_{16}, $C^{\#}_{17}$, and D_{18} are 287, 58, 195, 28, 1.5, 0.83, -28, -0.85, -2.63, 1.5, 40, 14, -49, 1, -3, 16, 24 and 28 cents.

Session 2pMUb

Musical Acoustics: Wind Instruments III and Mini Concert

D. Murray Campbell, Chair
School of Physics and Astronomy, Univ. of Edinburgh, Edinburgh EH9 3JZ, U.K.

Invited Paper

4:15

2pMUb1. Imitating the human voice: The renaissance cornett and sackbut. Murray Campbell (Sch. of Phys. and Astron., Univ. of Edinburgh, Edinburgh EH9 3JZ, U.K., d.m.campbell@ed.ac.uk)

In the 16th century the lip-reed instruments most commonly used in sacred and chamber music were the cornett (Italian "cornetto") and the sackbut (Italian "trombone"). The sound ideal described by contemporary writers for these instruments was clearly distinguished from the ceremonial and often military splendor of the trumpet ensemble: the cornett and the sackbut were expected to produce dynamic levels and timbres comparable to those of human singers. The acoustical features, which make it much easier to approach this ideal on the renaissance instruments than on modern trumpets and trombones are explored through measurements on original instruments and modern reproductions.

4:45—5:30 Mini Concert

Session 2pPA

Physical Acoustics and Biomedical Ultrasound/Bioresponse to Vibration: Numerical Methods for Weak Shock Propagation

Andrew A. Piacsek, Chair
Dept. of Physics, Central Washington Univ., Ellensburg, WA 98926-7422

Invited Papers

1:00

2pPA1. Numerical modeling of weak shock propagation: Past, present, and future. Andrew Piacsek (Dept. of Phys., Central Washington Univ., 400 E. University Way, Ellensburg, WA 98926, piacsek@cwu.edu)

This presentation provides an overview of the subject of numerical modeling of weak shock propagation, with the aim of providing context for subsequent talks in this session. First is a brief history of analytical approaches to the problem of weak shocks in both one and two dimensions, emphasizing the characteristic physical behavior of the solutions. This leads to a discussion of the physical processes that must be modeled, including the Rankine-Hugoniot conditions at the shock interface, nonlinear steepening, and self-refraction, refraction and diffraction due to an inhomogeneous medium, dissipation, relaxation, and heating of the medium. Next is a summary of paraxial numerical methods, specifically the NPE and KZK equations, and their recent applications, followed by a description of efforts to develop "wide-angle" versions of these codes. Computational challenges, and possible approaches, to extending these methods to three dimensions are also presented. Last, a brief introduction to new approaches to 3-D modeling will be given, along with a discussion of the applicability of computational improvements such as adaptive meshing and parallel processing.

1:20

2pPA2. Numerical simulation of acoustical shock waves: Beyond the parabolic approximation. Franck Dagrau, Mathieu Renier, François Coulouvrat, and Regis Marchiano (Univ. Pierre et Marie Curie, Institut Jean Le Rond d'Alembert (UMR CNRS 7107), 4 place Jussieu, 75005 Paris, France)

The KZK equation is the reference equation for the modeling of finite amplitude diffraction effects. It has been applied with success to many applications in various domains. Nevertheless, the parabolic approximation underlying the KZK equation limits the validity of this one to narrow-angle propagation. A new formulation of the Kuznetsov equation enables us to go beyond the parabolic approximation for the diffraction effects in the homogeneous case, the parabolic approximation being now limited only to the heterogeneous perturbation which is one order of magnitude smaller. This method formalizes and generalizes to the weakly heterogeneous case the

previous work of Christopher and Parker [J. Acoust. Soc. Am. **90**, 488–499 (1991)]. Special attention is paid to the implementation of boundary conditions (such as absorbing or perfectly matched layers). Several validation tests will illustrate the potential of the method, such as nonlinear focusing or scattering. Applications to sonic boom propagation in a turbulent atmosphere will finally be discussed.

1:40

2pPA3. Extension of the iterative nonlinear contrast source method to nonlinear media exhibiting tissue-like attenuation. Martin Verweij, Jacob Huijssen (Lab. of Electromagnetic Res., Fac. of Elec. Eng., Mathematics and Comput. Sci., Delft Univ. of Technol., Mekelweg 4, 2628 CD Delft, The Netherlands, m.d.verweij@tudelft.nl), and Nico de Jong (Erasmus Medical Ctr., Rotterdam, The Netherlands)

Recently, the Iterative Nonlinear Contrast Source (INCS) method has been developed to compute the pulsed acoustic pressure field in a nonlinear and lossless medium that extends over a very large, 3-D domain. This method solves the nonlinear wave equation by considering the nonlinear term as a distributed contrast source in a linear background medium. The full nonlinear wave field follows by convolving the background Green's function with successive estimates of the contrast source. It has been shown that by spatiotemporal filtering of the convolution factors, the method yields accurate field predictions even for a discretization approaching two points per smallest wavelength or period. The current paper discusses how attenuation, in particular tissue-like power law losses, may be incorporated into the method through the introduction of a causal compressibility relaxation function. The latter gives the background the desired loss behavior. Fields in nonlinear media with different power law losses are presented, as computed by the extended INCS method. Moreover, the ability of both the lossless and the extended INCS method to evaluate fields at the shock formation distance is shown. Especially the latter results show the significance of using a proper, tissue-like power law. [Work supported by STW and NCF.]

2:00

2pPA4. Weak shock propagation in the ocean and marine sediments. B. Edward McDonald (Acoust. Div. Naval Res. Lab, Washington DC 20375, mcdonald@ccs.nrl.navy.mil)

The nonlinear progressive wave equation (NPE) [B. E. McDonald and W. A. Kuperman, JASA 81, 1406 (1987)] was initially developed to model weak shock waves in a refracting medium (specifically, ocean acoustic convergence zones). It has been refined to include high-angle and two-way propagation. Inclusion of two-way propagation improves agreement with empirical power laws for peak pressure from an explosive as a function of range and explosive charge. The NPE is currently being used to examine weak shock propagation into marine sediments for studies related to mine countermeasures. An intriguing behavior results from the stress-strain relation for Hertzian granular media, in which stress is proportional to the 3/2 power of strain rate. Numerical and analytic solutions with the Hertzian nonlinearity reveal its most nonlinear behavior (shock formation) near zero stress, whereas in a fluid, shock formation results from high stress. Numerical experiments are presented to determine whether shocks resulting from Hertzian nonlinearity can be observed with nominal values of frequency-linear attenuation common to granular media. [Work supported by the Office of Naval Research.]

2:20

2pPA5. Modeling weak shocks produced by high-intensity focused ultrasound. Vera A. Khokhlova (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105/ Moscow State Univ., Moscow 119991, Russia), Olga V. Bessonova (Moscow State Univ., Moscow 119991, Russia), Michael S. Canney, Michael R. Bailey (Univ. of Washington, Seattle, WA 98105), Joshua E. Soneson (Food and Drug Administration, Silver Springs, MD 20993), and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

Shock waves of up to 100 MPa may form at the focus of high-intensity focused ultrasound (HIFU) transducers at clinically reported *in situ* intensities of up to 30,000 W/cm^2. The heating due to shocks is sufficient to boil tissue in milliseconds, which dramatically alters the treatment. Quantification of enhanced heating from shocks is therefore critical to treatment planning. In this work, several approaches and temporal grids of different resolutions were used to simulate HIFU fields. Peak positive pressure, which determines the shock amplitude, and thus the heating rate were found to be the most sensitive to the parameters of the numerical scheme. Heating rates calculated in modeling and estimated using weak shock theory from the measured and modeled waveforms were compared. Time to boil measured in tissue phantoms and tissue was used as a metric of the heating efficiency of shocks. It is shown that the bandwidth limitations in the waveform measurements result in underestimation of the heat rates, although boiling onset predicted in modeling agreed well with the experimental data. An experimentally validated numerical model thus can be an effective tool in both laboratory and clinical HIFU setting. [Work is supported by NIH EB007643 and NSBRI SMST01601.]

2:40

2pPA6. Numerical simulation of the propagation of three-dimensional helical shock waves in a weakly heterogeneous medium. Regis Marchiano, Francois Coulouvrat (Univ. Pierre et Marie Curie, Institut Jean Le Rond d'Alembert (UMR CNRS 7107), 4 place Jussieu, 75005 Paris, France), and Jean-Louis Thomas (Univ. Pierre et Marie Curie, 75005 Paris, France)

A numerical method for the simulation of 3-D acoustical shock wave propagation through a homogeneous or weakly heterogeneous medium is presented. It is based on a generalization of the KZ equation taking into account weak heterogeneities. The algorithm is based on a spectral treatment of the linear diffraction (angular spectrum), coupled with quasianalytical solutions for the heterogeneous part and for the nonlinear one. This last one solves Burgers' equation with the so-called Burgers-Hayes method using the potential instead of the pressure field. The combination of these several algorithms lead to the development of an efficient software, allowing to solve full 3-D problems for standard personal computers. This software is used to study the dynamics of helical shock waves also called acoustical vortices (AVs) which are the acoustical equivalent of optical vortices. The 3-D helical spatiotemporal wave field is characterized by azimuthal shocks. The dynamics of the so-called topological charge, an intrinsic property of AVs, is studied in the nonlinear regime through different focusing lenses.

3:00—3:20 Break

3:20

2pPA7. Numerical methods for weak shock propagation and diffraction around a microphone. Victor W. Sparrow (Penn State, Grad. Program in Acoust., 201 Appl. Sci. Bldg., University Park, PA 16802, vws1@psu.edu)

There are two interrelated goals for this presentation. The first goal is to give an overview of available numerical approaches for weak shock propagation. There are multiple approaches available, including formulations based on the Burgers equation, KZK or NPE equations, the second-order equations of nonlinear acoustics, the Navier-Stokes equations, or the Boltzmann equation. There are also multiple numerical techniques for discretizing each of underlying equations, and while some approaches give special treatment to weak shocks others handle the shocks automatically. The second goal of this presentation is to show recent simulation results for an explosion shock interacting in two dimensions with a microphone. The diffraction of the weak shock around the microphone housing does turn out to be amplitude dependent. It is only a modest effect, increasing for higher amplitudes. This knowledge is important for precise, close-range blast source characterization. In the future, the nonlinear diffraction of weak shocks around transducers also should be investigated for finite amplitude biomedical ultrasound, since the mechanisms for nonlinear diffraction are almost identical for ultrasound and explosion waves. [Work supported by U.S. Army ERDC-CERL.]

Contributed Papers

3:40

2pPA8. Simulations of converging shocks in water. Veronica Eliasson (Graduate Aerosp. Lab., Caltech, 1200 E. California Blvd., Pasadena, CA 91107, vero@caltech.edu), William D. Henshaw (LLNL, Livermore, CA 94551), and Paul E. Dimotakis (Caltech, Pasadena, CA 91125)

Numerical simulations of weak shocks in water traveling through a convergent geometry were performed. The convergent geometry is surrounded by an elastic solid, which is deformed by the fluid, thereby, generating elastic waves in the solid, which in turn affect the liquid, thus creating a coupled fluid-structure problem. Here, we use the Overture suit, which is a code for solving partial differential equations on curvilinear overlapping grids using adaptive mesh refinement. In particular, we use a multiphysics solver to solve the fluid-structure problem. The Euler equations with a stiffened equation of state are used in the fluid domain and the linear elasticity equations are used in the solid domain. The solutions at the interface between the fluid and the solid are matched using continuity of normal velocities and forces. Preliminary results indicate that the wave speed of the material has a large influence on the behavior of the converging shock. The numerical simulations are also compared to schlieren photographs and pressure measurements obtained from experiments. Results can enhance the design of marine structures with convergent sections subjected to dynamic loading events. [Work supported by ONR.]

3:55

2pPA9. Comparison of time and frequency domain approaches to simulate propagation of weak shocks. M.V. Averiyanov, P.V. Yuldashev (Faculty of Phys., Moscow State Univ., Moscow 119991, Russia, misha@acs366.phys.msu.ru), Ph. Blanc-Benon (LMFA, UMR CNRS 5509, Ecole Centrale de Lyon, 69134 Ecully Cedex, France 5509), and V.A. Khokhlova (Univ. of Washington, Seattle, WA 98105)

Simulations of acoustic fields using finite-difference methods are performed either in time or frequency domains. A method of fractional steps with an operator splitting procedure is frequently applied to solve nonlinear equations of the evolution type. Either time or frequency domain solvers can be used to calculate different terms in the equation over a propagation grid step. In this work, several algorithms that have been used to simulate quadratic nonlinear term in the Burgers or KZK-type evolution equations are applied to model the propagation of weak shocks. Shock capturing schemes of Godunov type, exact analytic solution with further extrapolation of the waveform over a uniform temporal grid, time-domain conservative schemes, direct modeling in the frequency domain, and asymptotic spectral approach are compared. The parameters of the schemes that would provide the results of the same accuracy, an artificial absorption necessary for stability of the schemes, resolution of shocks, and internal viscosity of the algorithms are discussed. It is shown that the Godunov-type algorithm is better suited to model weak shocks with sufficient accuracy achieved with only three temporal grid points per shock. [Work supported by the RFBR and NIH EB007643 grants.]

4:10

2pPA10. A shock-fitting method for general nonlinear progressive waves. Francois Coulouvrat (Univ. Pierre et Marie Curie, Institut Jean Le Rond d'Alembert (UMR CNRS 7107), 4 place Jussieu, 75005 Paris, France)

Many physical phenomena are concerned with the propagation of weak nonlinear waves that can be modeled under the form of a generalized Burgers equation. Physical examples include nonlinearities that can be either quadratic (nonlinear acoustical waves in fluids or longitudinal waves in solids), cubic (nonlinear shear waves in isotropic soft solids), or nonpolynomial (Buckley-Leverett equation for diphasic fluids, models for car traffic, Hertz contact in granular media). A new weak shock formulation of the generalized Burgers equation using an intermediate variable called "potential" is proposed. This formulation is a generalization to nonquadratic nonlinearities of the method originally proposed by Burgers himself in 1954 for his own equation, and later applied to sonic boom applications by Hayes et al. (1969). It is an elegant way to locate the position of a shock. Its numerical implementation is almost exact, except for an interpolation of Poisson's solution that can be performed at any order. It is also numerically efficient. As it is exact, a single iteration is sufficient to propagate the information at any distance. It automatically manages waveform distortion, formation of shock waves, and shock wave evolution and merging. The theoretical formulation and the principle of the algorithm are detailed and illustrated by various examples of applications.

4:25

2pPA11. A nonlinear computational method for the propagation of shock waves in ducts: Application to buzz-saw noise. Rasika Fernando, Yann Druon (Airbus France, Acoust. Environment Dept. Turbomachinery Acoust. (EDEA3), 316 route de Bayonne, 31060 Toulouse Cedex 03), Regis Marchiano, and Francois Coulouvrat (Univ. Pierre et Marie Curie, 75005 Paris, France)

A numerical method to compute shock wave propagation in uniform waveguides in 2-D and 3-D is presented. The solution is searched under the form of a modal solution of the Kuznetsov equation for which the modal amplitudes of the analytical linear modes are supposed to vary along the duct axis due to nonlinear interactions between the different modes and frequencies. This finally yields to a differential system on the mode amplitudes, which is solved numerically using a standard Runge-Kutta algorithm for ordinary differential equations after numerical truncation of the modal series. Examples are presented of the nonlinear evolution of the pressure field for a 2-D waveguide. One important 3-D application is the so-called "buzz-saw" noise occurring for high-bypass-ratio turbofan engines, when fan blade tip relative flows exceed the sound velocity. First results on 3-D simulations for nonlinear propagation of a saw-tooth waveform spiraling inside a hard-walled cylindrical will be presented. Extensions of the model to ducts with uniform flows or with liners (walls with impedance) will finally be discussed.

4:40

2pPA12. Evaluation of a wave vector frequency domain method for nonlinear wave propagation. Yun Jing and Greg Clement (Dept. of Radiology, Harvard Med. School, Brigham and Women's Hospital, Boston, MA 02155)

In this paper, a wave vector frequency domain method is presented to describe both forward and backward nonlinear wave propagations in acoustic media. A frequency domain nonlinear wave equation is first derived based on the Westervelt equation by taking the Fourier transform in the temporal and spatial domains. Since the model works directly in the frequency domain, arbitrary dispersion relations can be elegantly considered. The nonlinearity is further evaluated by multidimensional autoconvolution of the Fourier transform of the sound pressure. An implicit solution of the equation is proposed by employing the Green's function, and is solved numerically. Comparisons between the numerical simulations and the Burger's solution show good agreements within the shock formation distance. Error study is also carried out to test the efficiency of the model. It is found that the error grows as the initial pressure increases and the spatial resolution decreases. Finally, the present model is compared with a time-domain approach for a practical case, where the diameter of the transducer is 100 mm, the fundamental frequency is 1 MHz, and the focus range is 150 mm.

TUESDAY AFTERNOON, 19 MAY 2009 FORUM SUITE, 2:00 TO 4:40 P.M.

Session 2pSA

Structural Acoustics and Vibration: Emerging Applications of Structural Acoustics

Sean F. Wu, Chair
Dept. of Mechanical Engineering, Wayne State Univ., Detroit, MI 48202

Invited Papers

2:00

2pSA1. Intrinsic damping models and their role in predicting structural response. Allan D. Pierce (Dept. of Mech. Eng., Boston Univ., Boston, MA 02215, adp@bu.edu) and Adnan Akay (Bilkent Univ., Ankara, Turkey)

In many industries, it is often desired to estimate whether a structure can withstand severe broad-band high-frequency loads. Because the structures have many natural frequencies within the band of interest, the response is critically affected by the damping within the structure. One typically performs finite element model computations with the elastic modulus replaced by a complex number, the ratio of imaginary to real parts being the loss factor, this taken as frequency independent. The crudity of such a model invariably leads to designs that are either overly conservative or not inspiring confidence. The present paper revisits the topic of internal damping from a fundamental standpoint and adopts the viewpoints of [J. J. Thomson (1888), Clarence Zener (1948)], in which "all anelastic phenomena are regarded as the superposition of elementary processes in which the stress relaxes exponentially." The resulting general model involves a continuous distribution over relaxation times. Lumped parameter realizations require that the configuration of the structure be described by the positions of all the mass points, and also by those of a large number of "hidden variables" connected to the mass points by Maxwell elements (spring and dashpot is series). General implications of the model are derived and explained.

2:20

2pSA2. Measurement methodologies for the analysis of influence of surface panels and their vibrations on interior cabin noise. Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202, aa3199@wayne.edu) and Allan Pierce (Boston Univ., Boston, MA 02215)

In the design of quiet automobile compartments, one analyzes the influence of individual panels and enclosure surface leaks on noise received in a specific region (the location of the driver's ear position). Two recent papers [Hald *et al.*, Internoise (2006); Wolff, Internoise (2007)] suggest the needed measurement process can be shortened using a sound source at the typical listener position, taking measurements over the enclosure surface, and then invoking the principle of reciprocity. The present paper clarifies the analytical discussions of these earlier papers with improved and more nearly rigorous derivations and suggests improved measurement methodologies. One exemplary model considered is that of an arbitrarily shaped cavity with sound absorbing walls that are also moving due to external influences. The boundary condition on the pressure field inside the cavity is shown to be well approximated such that the normal component of the pressure gradient is a superposition of a constant times the pressure itself (locally reacting surface with finite impedance) plus a constant times the portion of the surface velocity that is due to the external sources. The pressure inside the cavity can be solved by Green's function techniques, and the Green's function can be proven to rigorously satisfy reciprocity.

2:40

2pSA3. Using structural acoustics methods to study new materials. J. D. Maynard (Dept. of Phys., The Pennsylvania State Univ., University Park, PA 16802)

In structural acoustics it is usually necessary to know the vibration properties of the structure under study, and this requires knowing the elastic constants of the materials involved. A relatively recent and powerful method for measuring elastic constants is resonant ultrasound spectroscopy (RUS). Very new and exotic materials for structures often require RUS; the reason is that new materials, when first fabricated, are often available only in very small samples (a few hundred micrometers) or as thin films on substrates, and RUS is the best method for handling such samples. On some occasions samples are fragile or chemically reactive so that they cannot be pol-

ished into the shapes required by conventional RUS. Thus, it has been necessary to develop a RUS analysis for arbitrarily shaped samples. This has been accomplished by borrowing a method from structural acoustics: the finite element method (FEM). The discussion of the application of FEM to RUS for arbitrarily shaped samples will include bridging the gap between theory (involving matrices as boldfaced symbols without indices in textbooks on FEM) and actual lines of computer code. Applications of the RUS method for materials will also be discussed.

3:00—3:20 Break

3:20

2pSA4. Spatial maps of modal damping from frequency response measurements. J. Gregory McDaniel, Craig Boucher, and Hande Öztürk (Mech. Eng. Dept., Boston Univ., 110 Cummington St., Boston, MA 02215)

A method is proposed for constructing spatial maps of modal damping for viscously damped structures. These maps are intended to be useful in assessing the effectiveness of spatially distributed damping, particularly in complex structures that are damped by a variety of mechanisms. The method uses frequency response measurements and modal analysis to determine the natural frequencies and mode shapes of the damped structure. The damping matrix is estimated directly from the frequency response measurements and is expressed as a sum over connectivity matrices weighted by dashpot constants. This sum is substituted into the well-known approximation for the modal loss factor that results by neglecting off-diagonal elements of the modally transformed damping matrix. The resulting expression quantifies the contribution of local dashpot properties to the modal loss factor, and may be visually presented as a spatial map of damping. A significant advantage of this approach is its independence from a detailed model of the structure or its damping, allowing it to be used for complex structures. The resolution of the map is directly correlating to the number of measurement locations, as will be illustrated by examples. [Work supported by ONR under Grant No. N000140810531.]

3:40

2pSA5. A three-dimensional Wiener-Hopf technique for general bodies of revolution, Part II: Substructuring and implementation. Rudolph Martinez and Carina Ting (CAA-Alion Corp., 84 Sherman St., Cambridge, MA 02140)

This paper begins with a brief review of Part I's main ideas and conclusions regarding: (a) Operator-induced symmetry and reciprocity leading to compatibility of the halfplanes of analyticity between the integral equation's non-translational kernel and those of the Fourier-transformed displacements and pressures on the surface of a scattering or vibrating body of revolution; and (b) subsequent spectral factorizations in juxtaposition to those of the classical Wiener–Hopf case for the planar geometry. The spatial and spectral domains in question refer to the arclength variable along the body's generator and to its transform wavenumber, respectively. The analysis described in this Part-II paper brings out some of the additional positive consequences of the pre-symmetrization of the integral equation that governs the response of the coupled fluid and structure. It then applies standard substructuring schemes to extend Part I's formulation to include departures from axisymmetry resulting from generic appendages as discontinuities along the shape's circumferential coordinate. The presentation ends with sample applications of its new 3-D Wiener–Hopf technique to canonical shapes of revolution, and with a summary of the physical insights made possible by the new approach relative to numerical results by boundary elements.

4:00

2pSA6. Surface wave testing of pavements. Nils Ryden (Eng. Geology, Faculty of Eng., Lund Univ., Box 118, SE-22100 Lund, Sweden)

Pavements are constructed using several layers of materials, and their durability depends on the quality of all of these strata. It is, therefore, valuable to be able to determine the properties of the layers nondestructively. A method is presented for evaluating the thickness and stiffness of multilayered pavement structures from guided waves measured at the surface. In this type of layered structure, the interaction of leaky Lamb waves in the embedded layers generates surface waves corresponding only to certain portions of the guided wave dispersion curves and branches measurable at the pavement surface. To resolve the different mode branches, the wavefield is measured at the surface by using a light hammer as the source and an accelerometer as receiver, generating a synthetic receiver array. The recorded data are transformed to a phase velocity spectrum, which is then inverted to give the layer properties using a global inversion algorithm. The theoretical background along with experimental results of the application to nondestructive testing of pavements will be presented. Ongoing research on noncontact air-coupled measurements is also demonstrated. This opens up the possibility for faster on-the-fly surface wave testing of pavement layers, since surface contact is no longer required.

4:20

2pSA7. Innovative structural acoustic strategies to reduce sound transmission through lightweight flexible structures. Donald Bliss, Qinxian He, Linda Franzoni, and Cassidy Palas (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708, dbb@duke.edu)

This paper describes research utilizing structural acoustics in novel ways to cancel the transmission of sound and vibration through multielement flexible barriers. Configurations analyzed include panels with different thickness and elastic modulus in different regions and layered structures connected by an elastic suspension. The purpose of the research is to demonstrate that flexibility and controlled resonant behavior can be used to block sound transmission even when structural damping is very low. Strategies are considered to alter vibrating surface wavenumber spectra to reduce coupling between the structure and the acoustic field. Another approach that can be employed is the utilization of structural wave cutoff with multielement multipath (MEMP) structures. Finally, multiple differentially tuned subsidiary elements acting as resonators can be used to greatly reduce the structural response. Examples of acoustic transmission

loss through panel barriers using different strategies are presented, and the potential advantages and possible shortcomings of various approaches are evaluated. Possible configurations for layered sound reduction materials are proposed. The work has particular application to the reduction of vehicle interior noise and addresses the need for good acoustic performance of the lighter weight flexible structures that will be used in the future.

TUESDAY AFTERNOON, 19 MAY 2009 EXECUTIVE SALON I/II/III, 2:00 TO 5:30 P.M.

Session 2pSC

Speech Communication and Psychological and Physiological Acoustics: Exploring the Relationship Between Cognitive Processes and Speech Perception (Lecture/Poster Session)

Amee P. Shah, Chair
Dept. of Speech and Hearing, Cleveland State Univ., Cleveland, OH 44115

Chair's Introduction—2:00

Invited Papers

2:05

2pSC1. Perceptual learning and expectations: Cognitive mechanisms in speech recognition. Howard Nusbaum (Dept. of Psych., The Univ. of Chicago, 5848 S. University Ave., Chicago, IL 60637, h-nusbaum@uchicago.edu)

There is substantial plasticity in adult perceptual processing of speech that cannot be accounted for by most theories. Even 1 h of training on low-intelligibility synthetic speech can improve recognition by 20 percentage points for novel words. Whether speech perception is based on auditory properties or articulatory properties, few theories acknowledge the role of cognitive processes, and some explicitly exclude these mechanisms. Evidence will be examined suggesting perceptual learning of phonetic information in adult listeners involves processes such as attention and working memory. In addition, evidence will be examined that speech perception can be influenced, if not typically guided, by listeners' expectations about speech. In some cases, expectations may be derived by explicit instruction to listeners, and in others, from contextual information. For example, research on talker normalization and perception of sinewave speech has demonstrated that listeners' expectations about the speech signal change perceptual processing. Evidence will be discussed suggesting mechanisms that mediate this effect of expectations and compare this to mechanisms involved in perceptual learning. Perceptual learning may depend on the same kinds of cognitive mechanisms involved in the top-down guidance of perception by expectations and these mechanisms may be necessary for stable phonetic perception. [Research supported by NIDCD.]

2:25

2pSC2. The role(s) of capacity limitations in speech perception. Alexander L. Francis (Speech, Lang., and Hearing Sci., Purdue Univ., 500 Oval Dr., West Lafayette, IN 47907, francisa@purdue.edu)

This paper reviews some current research exploring the role that capacity limitations play in recognizing speech, particularly in the presence of competing speech. Load theory [Lavie, Trends Cogn Sci **92** (2005)] suggests that perceptual capacity limits the number of objects (or features) that can be processed simultaneously, while cognitive capacity limits rejection of irrelevant information. Functional imaging studies show increased cognitive (working memory) load corresponding to failure of inhibition of (increased activity in) brain regions associated with processing distractor stimuli [de Fockert *et al.*, Science **291** (2001)], whereas increased perceptual load results in decreased activity in distractor-related areas [Rees *et al.*, Science **278** (1997)]. However, all of these studies were carried out in the visual modality, and it is unclear whether the same predictions were obtained in audition or speech perception. A series of experiments in which perceptual and working memory load were varied as listeners identified words produced by one talker while ignoring another provides qualified support for the predictions of load theory and lays the groundwork for subsequent studies of the effects of hearing impairment, cognitive deficits, and perceptual and cognitive training on the perception of speech in competing speech. [Work supported by NIDCD R03DC006811.]

2:45

2pSC3. Hearing the forest despite the trees: Perceptual learning of systematic variation in speech. Lynne C. Nygaard (Dept. of Psych., Emory Univ., Atlanta, GA 30322)

A fundamental problem for the understanding of spoken language processing is listeners' robust perceptual constancy in the face of enormous variability in the instantiation of linguistic form. On the one hand, listeners are sensitive to the fine-grained structure of linguistic segments that signal differences among talkers and speaking styles. On the other hand, listeners tolerate large discontinuities in this same fine structure forming robust, perceptually constant linguistic categories. Data from a perceptual learning paradigm addressing both the limits and flexibility of speech perceptual mechanisms will be presented and discussed in light of evidence for perceptual learning of accented speech. In a series of studies, listeners were exposed to accented English under learning conditions in which

the opportunity to compare across instances varied. At test, generalization and long-term retention of perceptual learning were evaluated. The results suggest that adaptation to lawful variation in speech is fundamentally influenced by the character and structure of the learning experience. These findings suggest cognitive constraints on behavioral and representational plasticity in speech perception and spoken language processing.

3:05

2pSC4. The nature and time course of the relationship between surface variability and linguistic processing. Amee Shah (Dept. of Health Sci., Cleveland State Univ., 2121 Euclid Ave., MC 429, Cleveland, OH 44115) and Conor McLennan (Cleveland State Univ., Cleveland, OH 44115)

It is becoming increasingly prevalent to study and understand the role of surface information in the process of speech perception and to attempt to accommodate these variables in theories of speech perception. Indeed, theories of speech perception increasingly require an account of the precise circumstances under which surface details affect the ease with which listeners access spoken words. To date, the majority of studies have focused on manipulating various components of surface information (e.g., speaking rate, talker differences, dialect difference, and so on) to determine whether these manipulations matter to the processing task. The present set of studies in our laboratory represents the converse component of this relationship in order to determine whether linguistic complexity can affect or modulate listeners' overt subjective impressions of the surface information. Two different methodologies, including semantic priming and repetition priming, were used manipulate ease of lexical access and determine the role on listeners perception of surface details (foreign accent). Additionally, the time course of the effect of linguistic context on accent rating judgments was studied as it manifested itself differently at different points during perceptual processing. Overall, these studies help understand the complex nature of the relationship between surface variability and linguistic processing.

3:25

2pSC5. Influences of auditory object formation on speech perception. Barbara G. Shinn-Cunningham and Dali Wang (Boston Univ. Hearing Res. Ctr., Auditory Neurosci. Lab., 677 Beacon St., Boston, MA 02215)

In common social settings, our ears receive a jumbled mixture of sound coming from different sources in different directions. A normal hearing listener is adept at using the low-level acoustic structure in the sound mixture to organize this cacophony into auditory objects, each corresponding to a distinct talker (i.e., performing auditory scene analysis). Typical listeners are also proficient at focusing on the talker of interest and ignoring distracting objects (i.e., directing selective attention to the desired object). Often, despite the efficacy of object formation and selective attention, portions of an attended object are inaudible, masked by sources that overlap the attended source in time and frequency. In such situations, listeners "fill in" the missing speech and extract a meaningful, whole message from the sound mixture, a process known as "phonemic restoration." This talk will review some recent studies demonstrating that the process of phonemic restoration, which relies on high-level, acquired knowledge about speech structure and meaning, is directly impacted by low-level acoustic attributes that influence auditory grouping. These results demonstrate the inter-connected nature of the high-level processes that allow listeners to communicate in challenging settings.

3:45—3:55 Break

3:55—4:25 Panel Discussion

4:25—4:30 Break

Contributed Papers

All posters will be on display from 4:30 p.m. to 5:30 p.m. To allow all contributors an opportunity to see other posters, contributors of odd numbered papers will be at their posters from 4:30 p.m. to 5:00 p.m. and contributors of even numbered papers will be at their posters from 5:00 p.m. to 5:30 p.m.

2pSC6. The learning of lexical tones. Sunjing Ji (Dept. of Linguist., Univ. of Arizona, Douglass Bldg., Rm. 200E, Tucson, AZ 85719, sunjing@email.arizona.edu)

Our understanding of speech perception suggests that categorization is an essential task of our cognitive system. Based on previous research on segment perception, this study explores the questions of what mechanism is involved in the acquisition of lexical tones and whether contour tones are decompositional in the process of cognitive learning. An experiment adopting the statistical training method from Maye *et al.* (2002) and Hayes-Harb (2007) was conducted on the learning of lexical tones by adult naive tone learners. The results show that statistical distribution of stimuli along a pitch continuum in training affects naive tone learners' formation of level lexical tone categories. In addition, the perception of level tones behaves more like the perception of vowels rather than consonants. Moreover, exposure to statistical training on a level tone continuum alone affects participants' performance in contour tone discrimination. This indicates that participants were able to apply level tone categories to generate new mental categories of contour tones, providing psychological evidence for the auto-segmental argument that contour tones are composed of a sequence of multiple level tones.

2pSC7. Effects of differences in fundamental frequency on cross-formant grouping in speech perception. Robert J. Summers, Brian Roberts (Psych., Sch. of Life and Health Sci., Aston Univ., Birmingham B4 7ET, UK, r.j.summers@aston.ac.uk), and Peter J. Bailey (Dept. of Psych., Univ. of York, Heslington, York YO10 5DD, UK)

In isolated syllables cross-formant perceptual grouping can be promoted by a common fundamental frequency (F0), but this effect is relatively weak [Darwin, Q., J. Exp. Psychol. **33A**, 185–207 (1981)]. Few studies have explored the role of F0 in cross-formant grouping using sentences. Three-formant (F1+F2+F3) analog of almost continuously voiced natural utterances were synthesized using a monotonous glottal source (F0=150 Hz). Perceptual organization was probed by presenting stimuli dichotically (F1+F2C+F3; F2), where F2C is a competitor for F2 that listeners must resist to optimize recognition. Competitors were created using time-reversed

amplitude and frequency contours of F2, and F0 was manipulated ($\Delta F0 = \pm 8, \pm 2$, or 0 semitones relative to the other formants). Adding F2C reduced intelligibility for most utterances, and this reduction was greatest overall when $\Delta F0 = 0$. However, as for isolated syllables, the effect of $\Delta F0$ was quite small. Competitor efficacy was also somewhat dependent on absolute F0. Manipulation of relative F2C level indicated that competitor efficacy was not due primarily to energetic masking. The modest effect of $\Delta F0$, both for sentences and isolated syllables, suggests that cross-formant grouping depends more on other factors, for example common patterns of modulation. [Supported by EPSRC]

2pSC8. The perceptual organization of sine-wave speech under competitive conditions. Brian Roberts, Robert J. Summers (Psych., Sch. of Life and Health Sci., Aston Univ., Birmingham B4 7ET, UK, b.roberts@aston.ac.uk), and Peter J. Bailey (Univ. of York, Heslington, York YO10 5DD, UK)

Speech comprises dynamic and heterogeneous acoustic elements yet is heard as a single perceptual stream, even when accompanied by other sounds. The relative contributions of grouping primitives and of speech-specific factors, for example modulation patterns, to the perceptual coherence of speech are unclear, and the critical acoustical correlates of the latter remain unknown. The parametric manipulations possible with simplified speech, such as sine-wave analogs, make them attractive experimental stimuli to explore this issue. Given that the factors governing perceptual organization are generally revealed only where competition operates, the second-formant competitor (F2C) paradigm was used, in which the listener must resist competition to optimize recognition [Remez et al., Psychol. Rev. 101, 129–156 (1994)]. Three-formant (F1+F2+F3) sine-wave analogs were derived from natural utterances and presented in a dichotic configuration (one ear = F1+F2C+F3; opposite ear = F2). Different versions of F2C derived from F2 by manipulating its amplitude and frequency contours were effective at reducing intelligibility except when the frequency contour was constant. Manipulation of relative F2C level indicated that competitor efficacy was not due primarily to energetic masking. The findings suggest that modulation of the frequency, but not the amplitude, contour is critical for cross-formant grouping. [Work supported by EPSRC.]

2pSC9. Developing audio-visual associations in infancy. Francisco Lacerda, Eeva Klitfors, and Ellen Marklund (Dept. of Linguist., Stockholm Univ., SE-106 91 Stockholm, Sweden, frasse@ling.su.se)

To investigate how possible biases in matching of object sizes with sound intensities may be influenced by the ecological relevance of the potential associations, three groups of Swedish infants (4, 6, and 8 months of age) are being tested in their preferences to match speech or nonspeech sounds of different intensities with visual stimuli showing faces or unanimated objects of different sizes. The infants' audio-visual matching preferences are measured by the cumulative time that the infants' gaze vector was directed to each of the faces and objects simultaneously displayed on the screen, in response to the presentation of an acoustic stimulus (either a speech sound or a nonspeech sound). The results appear to suggest a general trend towards matching of visual prominence with sound intensity, when the ecological relevance of the audio-visual associations was high (i.e., matching speech sounds with faces and nonspeech sounds with unanimated objects) but the pattern seems to be disturbed when only nonecologically relevant matches are possible. [Work carried out within grants from The Bank of Sweden Tercentenary Foundation, Grant No. K2003:0867, MILLE, from EU-NEST Project No. 5010, CONTACT, and from the Knut and Alice Wallenberg Foundation, Grant No. KAW 2005.0115.]

2pSC10. Voice familiarity helps infants tackle variability in the speech signal. Marieke van Heugten and Elizabeth Johnson (Dept. of Psych., Univ. of Toronto Mississauga, 3359 Mississauga Rd. N., Mississauga, ON, Canada, L5L 1C6, marieke.vanheugten@utoronto.ca)

The acoustic realization of lexical items produced by different speakers can vary greatly. Current research suggests that infants, unlike adults, struggle to cope with this lack of invariance in the realization of words. Although 7.5-month olds are able to recognize words across different utterances when produced by speakers of the same gender with similar voices, they fail to do so when target words produced in a female voice are subsequently produced in a male voice [Houston and Jusczyk (2000)]. Note that all work in this area has used disembodied unfamiliar voices to test infants. In the current study, we ask whether infants might perform better under more ecologically valid conditions, i.e., when tested on familiar rather than unfamiliar voices. Using the headturn preference procedure, infants were familiarized with passages spoken by their mother. During the test phase, they were presented with their father's voice producing isolated repetitions of familiarized target words. Preliminary results suggest that infants may recognize words across different utterances produced by speakers of different genders if they are highly familiar with both the male and female speakers. In other words, infants may handle variability in the realization of words better when tested on familiar rather than unfamiliar voices.

2pSC11. The development of infants' sensitivity to modified spectral tilt: Fricatives, approximants, and vowels. Elizabeth Beach and Christine Kitamura (MARCS Auditory Labs., UWS, Bdg. 5 Bankstown Campus, Locked Bag 1797, Penrith South DC 1797, NSW, Australia)

In early infancy, speech perception is based on innate psychoacoustic thresholds allowing young infants to discriminate a wide range of speech contrasts. However, as infants accumulate knowledge of their native language, they begin attuning to native speech sounds: first vowels around 6 months; then consonants around 9–12 months. Now that hearing-impaired infants are being diagnosed and fitted with hearing aids early in life, there is a need to investigate how speech is amplified in infant hearing aids. This study examined whether infants would benefit from positive, negative, or unmodified spectral tilt. Spectral tilts of +6dB/octave and −6dB/octave (as found in hearing aids) were applied to three speech contrasts (fricatives, approximants, vowels) and in a third condition the contrasts remained unmodified. Normal-hearing 6- and 9-month-olds were tested using an habituation procedure to determine whether positive, negative, or unmodified tilt aids discrimination of the speech contrasts. The results showed 6-month-olds benefit when relevant frequency information is amplified, i.e., high-frequency for fricatives and approximants, low-frequency for vowels. Nine-month-olds, on the other hand, discriminated all contrasts best when unmodified. The developmental change between 6 and 9 months might be explained by infants shifting from an acoustic to a linguistic mode of speech perception.

Session 2pSP

Signal Processing in Acoustics and Underwater Acoustics: Detection and Classification of Underwater Targets II

Patrick J. Loughlin, Cochair
Dept. of Bioengineering, Univ. of Pittsburgh, Pittsburgh, PA 15261

Jack McLaughlin, Cochair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105

Invited Papers

1:30

2pSP1. Alternative signal processing models for broadband underwater propeller sounds. Les Atlas, Pascal Clark (Dept. of Elec. Eng., Univ. of Washington, Campus Box 352500, Seattle, WA 98195, atlas@u.washington.edu), and Ivars Kirsteins (Naval Undersea Warfare Ctr., Newport, RI)

Underwater propeller sounds often have audibly rhythmic characteristics. These sounds are commonly modeled as a broadband noise carrier multiplied by a rhythmic modulator, where analysis of the modulator wave form can provide information relating to the speed and identity of the source vessel. A natural assumption is to posit a purely periodic modulator, considering the rotary action of the propeller, which is useful for deriving maximum-likelihood estimates of the shaft rate [Lourens and du Preez, J. Ocean. Eng. **23**, 4, (1998)]. Indeed, recorded propeller data are unmistakably rhythmic, but the assumption of a strictly periodic or cyclical modulator wave form is questionable. A critical analysis of the periodicity assumption is presented and compared to three alternatives. Maintaining the product form in which the modulator wave form multiplies a random broadband carrier, the possible modulator signals are generalized to include: (1) almost-periodic functions expressible by nonharmonic trigonometric series; (2) multiplicative cascades of possibly complex-valued modulator functions; and (3) stochastic processes in which random state transitions constitute a sense of repeatability or rhythm. In this work, three modulator models are contrasted as estimators of periodicity, using actual propeller sound recordings to validate these estimates. [This work is supported by the Office of Naval Research.]

1:50

2pSP2. The propagation of noise in a dispersive medium. Leon Cohen (Dept. of Phys., City Univ. of NY, Hunter Coll., 695 Park Ave., New York, 10021)

If noise is generated at a particular region, it will generally propagate and be nonstationary in position and time. We present a phase space approach to understand the propagation of noise in a dispersive medium and to present methods to ascertain when the noise is stationary or quasi-stationary. The damping case is also discussed. In addition, we present approximate methods that allow one to study the statistical properties of the evolution of the noise field in space and time in a relatively simple way. A number of examples will be presented. [Work supported by ONR.]

2:10

2pSP3. Classification of shallow water passive sonar signals using stochastic channel model. Maya Gupta, Hyrum Anderson, and William Mortensen (Dept. of Elec. Eng., Univ. of Washington, Seattle, WA 98195)

This paper addresses the problem of classifying passive sonar signals propagating in shallow water channels. A key problem is that such signals are corrupted by noise and multipath, which we model stochastically in terms of the expected multipath and noise, and the variance of the multipath and noise. We assume that free-field (e.g., deep water) training signals are available. We show that for classification, the accuracy can be improved by not committing to an estimate of the channel or signal, but rather by marginalizing out the channel and noise uncertainty. Specifically, we propose variants of the quadratic discriminant analysis (QDA) classifier and the support vector machine classifier (SVM) that probabilistically account for the unknown channel effects, and which avoid ill-posed deconvolution. The proposed classifiers can work either directly on a time signal or on subband power features. Results on simulated data and real Bowhead whale vocalizations show that we can significantly improve classification performance over traditional methods over a range of signal-to-noise ratios. In addition, we will discuss the case where the classification decision can be made by multiple cooperating receivers, such as in a sonar network.

2:30

2pSP4. Finding physically motivated classification features using finite element models. Jack McLaughlin and Lane M.D. Owsley (Univ. of Washington, Appl. Phys. Lab., Box 355640, 1013 NE 40th St., Seattle, WA 98105)

Finite element models (FEMs) for simple shapes such as spheres, cylinders and pipes can be used to uncover potential features for classification by sonar systems. Analyses were conducted of the simulated, bistatic returns over a wide range of aspects using FEMs for a solid steel cylinder and a cement pipe all in free space. Salient characteristics noted in this way were then sought in field data collected

from the same or similar objects during the SAX04 experiment and two other data collections conducted in a test pond at the Naval Surface Warfare Center in Panama City, Florida. Characteristics robust enough to appear in both simulated data and field data are candidates to serve as classification features for similarly shaped objects. It is found that FEMs can provide important pointers to classification features that are closely tied to the physics of reflection. As such, the variability of these features in the face of object burial and multipath can be more easily assessed than statistical spectral features. [This work was supported by the U.S. Navy Office of Naval Research.]

2:50

2pSP5. Kirchhoff scattering from underwater targets modeled as an assembly of triangular facets. Ahmad T. Abawi (HLS Res., Inc., La Jolla, CA)

Scattering from underwater targets is a challenging problem, mainly because many of the techniques developed in scattering theory are for radar applications and, therefore, appropriate only in free space. Underwater scattering occurs in a waveguide, which in most cases requires a modification if not a reformulation of the methods developed for scattering from targets in free space. The most accurate solution of scattering from a target in a waveguide can be obtained by solving the wave equation in an environment that contains the waveguide and the target and simultaneously imposing boundary conditions on the boundaries of the waveguide and on the surface of the target. However, this technique is numerically intensive making it impractical in most cases of interest. As a means of offering a more practical solution, we show how to solve the scattering problem from a target in a waveguide by (1) modeling it as an assembly of triangular facets, (2) computing scattering from each facet analytically using the Kirchhoff approximation for each set of incident and scattered plane waves, and (3) combining the solutions coherently. We compare this solution with those obtained using more accurate methods such as the virtual source technique.

3:10—3:20 Break

Contributed Papers

3:20

2pSP6. Change detection deconfliction process for sonar clutter items. John Dubberley and Marlin Gendron (Naval Research Lab. Code 7441, Stennis Space Ctr., MS 39529, john.dubberley@nrlssc.navy.mil)

When resurveying a seafloor area of interest during change detection operations, an automated method to match found bottom objects with objects detected in a previous survey allows the surveyor to quickly sort new objects from old objects. Here we will demonstrate a software system that accomplishes change detection. The change detection system contains modules for automatic object detection by geospatial bitmap technique, object collocation, feature matching using shadow outlining, scene matching by control point matching, and visualization and filtering capabilities. Emphasis will be placed on the new elements of the system, namely, shadow outlining and optional spatial filtering.

3:35

2pSP7. Processing techniques applied to underwater target echoes and their relationships. Brian H. Houston (Naval Res. Lab., Code 7130, 4555 Overlook Ave., Washington, DC 20375-5320, brian.houston@nrl.navy.mil), Joseph A. Bucaro (Naval Res. Lab., Washington, DC 20375-5320), Larry Kraus (Global Strategies (North America), Crofton, MD 21114), Harry J. Simpson (Naval Res. Lab., Washington, DC 20375-5320), and Timothy J. Yoder (Global Strategies (North America), Crofton, MD 21114)

Interest in acoustic scattering from underwater unexploded ordnance (UXO) has been increasing because of the growing need for sonar technology able to detect submerged UXO and to efficiently separate these detections from those due to natural and man-made clutter. Recent efforts range from laboratory and numerical studies seeking to understand the basic structural acoustic echo formation processes and the environmental effects on them to field exercises determining how well a particular detection /identification technology performs. This paper discusses several important scattering-based constructs, which have been utilized in both the basic UXO studies and the technology exercises as well. They include: (1) $\sigma(\omega, \theta)$, the frequency/angle-dependent scattering cross section; (2) standard imagery, oftentimes implemented with synthetic aperture arrays; (3) reflection tomography imaging; and (4) supersonic imaging using holographic techniques. Without a clear appreciation for the differences among the constructs, it is difficult to properly relate the results of the various studies. We have measured the broadband scattering from several underwater UXO targets using laboratory, rail-based, and AUV systems and processed the data using the above constructs. The use of a common, high quality UXO scattering data base in this process helps clarify the differences and relationships among the constructs. [Work supported by SERDP and ONR.]

3:50

2pSP8. Bistatic synthetic aperture sonar images of penetrable cylinders. Christopher Dudley and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814)

Bistatic synthetic aperture sonar (SAS) and acoustic holographic images and monostatic backscattering by penetrable tilted cylinders in water were investigated. Refracted rays internally reflected from the cylinder's backside often dominated the scattering. Ray theoretic models predict scattering enhancements for cylinders made of some isotropic and anisotropic materials. Experiments involved solid cylinders and liquid-filled cylindrical shells of polymers and fiberglass. The experiments used transient insonification and ka in the range of $9<ka<40$, where ka is a product of the acoustic wave number and the cylinder radius. Bistatic measurements used a hydrophone scanned along a line. Transversely isotropic fiberglass rods produced rich features in the images and in the backscattering plotted as a function of tilt and time. The timing of some features could be explained from a ray model based on properties of quasi-shear and quasi-longitudinal waves computed from the stiffness matrix measured for the fiberglass. Bright features for solid cylinders and liquid-filled shells were typically associated with refracted waves having phase velocities less than the speed of sound in water. For some tilt angles, radiated wavefronts were relatively flat producing enhanced farfield scattering. [Work supported by ONR.]

4:05

2pSP9. Bistatic synthetic aperture sonar and acoustic holographic imaging of a tilted circular elastic plate. Neil Tuazon and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814, marston@wsu.edu)

Circular plates have axial symmetry and are sometimes used as components of items relevant to applications of sonar to detection and classification. Measurements of the backscattering spectrum as a function of tilt angle for an isolated circular plate reveal bright features associated with the excitation of compressional and flexural waves on the plate. Some of the features were previously found relevant to the interpretation of acoustic holographic images based on 2-D hydrophone scans close to a circular plate [B. T. Hefner and P. L. Marston, ARLO **2**(1), 55–60 (2001)]. In the present research, bistatic synthetic aperture sonar (SAS) and holographic images were constructed from line scans of a hydrophone at a large distance from

the plate using a method previously summarized [K. Baik, C. Dudley, and P. L. Marston, J. Acoust. Soc. Am. **121**, 3203 (2007)]. Because of elastic responses, the images are bright even when the receiver scan line does not intersect specular reflections from the plate. The radiation by the transient elastic waves excited on the plate can significantly increase the contrast of SAS and holographic images. [Work supported by ONR.]

4:20

2pSP10. Optical simulation of bistatic specular reflection by a rigid cone on a flat surface. Anthony R. Smith, Matthew D'Asaro, and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814, asmith8@mail.wsu.edu)

To gain insight into the high-frequency scattering of sound by cone-shaped objects near flat reflecting surfaces, a small conical mirror was placed on a flat horizontal mirror and illuminated by a tilted laser beam. The symmetry axis of the cone was vertical. Light scattering patterns were recorded by imaging with a camera on a white vertical screen that intersected the scattering. Bright regions on the screen were associated with specular reflections from the cone and from the flat mirror. To determine which contributions depended on the presence of the flat mirror, the screen was also imaged with the cone on a nonreflecting surface. The patterns were recorded with the screen placed sideways relative to the laser beam as well as in front of the cone and behind the cone. The cone half-angle was 45 deg but the grazing angle of the laser beam could be changed. The bright patterns were described by a geometric analysis of bistatic reflection previously summarized [P. Marston, J. Acoust. Soc. Am. **124**, 2584 (2008)] modified so as to allow for images caused by the flat mirror. The pattern for a vertical circular cylinder on a mirror was also studied. [Research supported by ONR.]

4:35

2pSP11. Modelled Doppler and range resolutions of active pulses in simulated environments. Andrew Holden (Dstl Farnborough, Ively Rd., Farnborough, Hampshire, GU14 0LX, UK)

The ability of an active sonar waveform to simultaneously estimate the range and velocity resolution of a target can be predicted by the wideband ambiguity function. Although this gives good insight into the capabilities of different pulse types, it does not replicate a real active sonar where a transmitted pulse is propagated through an environment and the received pulse is cross-correlated with a number of frequency and time shifted replicas of the transmitted pulse. In order to predict this a ray trace model is used to calculate the travel time and received amplitude of each ray path from the transmitting sonar to a target and back to include seabed and sea surface scattering where complex broadband pulse types can be modelled using a sub-band approach. This allows the determination of Doppler and range resolution for a variety of pulse types including pulse trains of LFM pulses in simulated environments, especially highly reverberate shallow water environments.

TUESDAY AFTERNOON, 19 MAY 2009 PAVILION WEST, 1:00 TO 5:05 P.M.

Session 2pUW

Underwater Acoustics and Acoustical Oceanography: Session in Honor of Ralph Goodman and His Contributions to the Acoustics of Bubbles and Other Works

Jerald W. Caruthers, Cochair
Dept. of Marine Sciences, Univ. of Southern Mississippi, Stennis Space Center, MS 39529

Kenneth E. Gilbert, Cochair
National Ctr. for Physical Acoustics, Univ. of Mississippi, University, MS 38677

Steve Stanic, Cochair
Naval Research Lab., Stennis Space Center, MS 39529

Chair's Introduction—1:00

Invited Papers

1:05

2pUW1. Ralph Goodman, builder of laboratories. Samuel W. Marshall (P.O. Box 668, White Stone, VA 22578, swmiii@yahoo.com)

Ralph had an extraordinarily successful career in building programs and laboratories. He started with an overall vision of what was to be built while determining the principle obstructions. He carefully avoided early details which would most often be overtaken by events. In addition, he was an eminent scientist whose acumen in a number of technical areas permitted him breadth of very constructive technical leadership. Besides a history of how Ralph got into underwater acoustics and his first major position in management, some personal anecdotes will be presented to emphasize how he successfully dealt with people and with some very difficult situations in this and subsequent undertakings.

1:25

2pUW2. The Italian connection: Ralph Goodman's close ties to NATO over 45 years. Finn B. Jensen (NATO Undersea Res. Ctr., 19126 La Spezia, Italy, jensen@nurc.nato.int)

Ralph's enduring and very intense professional relationship with SACLANTCEN (now NURC) in La Spezia, Italy, started in September 1961 when he joined the Centre as a young scientist. He worked in Italy for two years publishing on "convergent zone propagation" and "reverberation from smooth ocean floors." This was also the time of his first sea trial, which on Ralph's watch nearly led to the ship running aground. His love for life in Italy and the NATO Centre as such led to many visits in the coming years. He returned in 1981 to become the first American director of SACLANCEN. His primary task was to plan and undertake the construction of a new NATO research vessel named Alliance, a task which involved many crucial decisions about ship specifications and performance, plus guidance to the in-house ship building team, and excruciating negotiations with NATO agencies to secure the funding. Ralph's broad scientific background and superb people management skills ensured the success, and his "yacht" is still one of the best research vessels around, here 20 years later. Ralph's tenure as director lasted an unprecedented 6 years and he left in 1987 having secured SACLANTCEN a top-rated seagoing capability for years to come.

1:45

2pUW3. Coupled ocean-acoustics studies at Navy and NATO laboratories: The legacy of Ralph Goodman. Steve A. Piacsek, Charlie N. Barron (Oceanogr. Div., Naval Res. Lab., Stennis Space Ctr., MS 39529, piacsek@nrlssc.navy.mil), and Michael Porter (HLS, Inc., La Jolla, CA 92037)

A brief history of collaborations between ocean and acoustic modelers is given under the directorships of Ralph Goodman at NORDA and SACLANT, and of continuing coupled modeling studies at NRL built on previous studies he sponsored. At NORDA the Numerical Modeling Division consisted of an acoustic and an ocean branch, one of the first known instances of such close-knit administrative units for ocean and acoustic modelers. At SACLANT, close collaborations were strongly encouraged by Ralph Goodman and the respective group leaders. The results of these studies eventually appeared among the first published papers on the effects of mixed layer and Gulf Stream properties on surface-duct propagation (*Computational Acoustics: Ocean-Acoustic Models and Supercomputing* (North Holland Press, 1990); *Oceanography and Acoustics: Prediction and Propagation Models* (AIP Press, 1994). At NRL, these studies were restarted recently: evaluation of sonic layer depth relative to mixed layer depth [JGR **113** (2008)]; the acoustic impact on marine mammals [AGU Ocean Sciences Meeting (2006)]; and the acoustic impact of propagating internal solitons [IEEE Journal of Ocean Engineering (2007)]. Examples will be given from some of these studies to show progress in the field.

2:05

2pUW4. Ralph Goodman and Hamiltonian acoustics. W. A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., UC San Diego, La Jolla, CA 92093-0238)

Much of the underwater acoustic community that has worked with ray theory starts from the eikonal-ray equations. Since these equations, which are solved numerically, can be derived from a high-frequency asymptotic expansion, Fermat's/Hamilton's principle is typically bypassed. Historically, the introduction of the Hamiltonian formulation of underwater acoustics ray has come from scientists with a background in classical physics and/or from researchers whose goals were to study complicated environments, chaos, etc., using the well developed formalism of Hamilton's equations and perturbation theory. Mostly, these latter applications go back to the 1980s. Among Ralph Goodman's earliest work in underwater acoustics is a paper [R. Goodman and R. B. Duykers, J. Acoust. Soc. Am. 34, 960–962 (1961)] in which he derived the Hamiltonian for convergence zone rays in the harmonic oscillator approximation and then solved the equations by inspection. This paper must be one of the earliest applications of the Hamiltonian formulation to underwater ray acoustics and a part of the beginning of an illustrious career in the underwater acoustics community.

2:25

2pUW5. Bubble research with Ralph Goodman at Penn State. Kenneth E. Gilbert (Natl. Ctr. for Phys. Acoust., Univ. of Mississippi, University, MS 38677)

During the years 1992–1998 at Penn State, I collaborated closely with Ralph Goodman on scattering from the near-surface oceanic bubble layer. Ralph's enthusiasm, physical insight, and encyclopedic knowledge of at-sea experiments proved to be invaluable. I will outline the research and discuss Ralph's contributions to graduate theses and scientific papers that were published. I will also give some personal observations on what it was like collaborating with Ralph.

2:45

2pUW6. Extending bubble measuremenets below 20 μm: in memory of Ralph Goodman's contributions. Helen Czerski and David Farmer (The Graduate Sch. of Oceanogr., Univ. of Rhode Island, S Ferry Rd., Narragansett, RI 02882)

No matter which aspect of underwater acoustics we worked on, conversations with Ralph Goodman invariably added insight and often motivated new lines of investigation. This was certainly true of our interest in measuring bubble size distributions. We discuss the extension of acoustical measurement of bubbles to radii below 20 μm, a topic of particular relevance to the study of optical scatter in the ocean. Extending the frequency range of a resonator to 1 Mhz would in principle allow measurement down to ~3 μm, but significant challenges arise due to the confounding effects of geometric scatter from larger bubbles. This problem becomes greater for measurements of smaller bubbles, requiring a careful analysis of the accuracy limits imposed on the inversion procedure. Preliminary steps along this path are discussed with examples drawn from recent field measurements.

3:05—3:20 Break

3:20

2pUW7. Anomalous dispersion and pulse propagation in oceanic bubble clouds. Frank S. Henyey (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, frank@apl.washington.edu)

During a planning meeting for the Scripps pier bubble experiment, Ralph Goodman suggested we apply the Kramers-Kronig relations to bubble cloud data. By "we," I assumed he meant that I should do it. I applied the Kramers-Kronig relations to David Farmer's Scripps pier data from his modified Medwin resonators. This allowed a determination on which data had lower accuracy than the rest. Using the high accuracy data, I constructed a model of the complex sound speed. No formula relating bubble size distributions was needed. This model has anomalous dispersion in a frequency band that nearly coincides with a region in which the group speed is higher than the sound speed in bubble-free water. Summerfeld's theorem prohibits any signal from traveling that fast. The consequences for pulse propagation within bubble clouds was determined, using pulses having the same frequency spectrum as transmitted by Jerry Caruthers. The predicted nature of the received pulse is only partially described by Brillouin's studies, done before computers existed.

3:40

2pUW8. Conversations with Ralph about bubbles in the surf zone. Grant B. Deane (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238, gdeane@ucsd.edu)

Ralph Goodman made many contributions to the field of oceanic bubbles. This talk will focus on conversations with Ralph about bubbles in the surf zone. Questions that Ralph asked about surf zone bubbles included: where are they, and what are their acoustical properties? These fundamental questions will be examined in the light of Ralph's own experimental and theoretical work, with some contributions from the author. [Work supported by the Office of Naval Research.]

4:00

2pUW9. Ralph Goodman's recent contributions on oceanic bubbles. Jerald W. Caruthers (Dept. of Marine Sci. Univ. of Southern Mississippi, 1020 Balch Blvd., Stennis Space Ctr., MS 39529, jerald.caruthers@usm.edu) and Steve Stanic (Naval Res. Lab., Stennis Space Ctr., MS 39529)

Many of us are never too old to have an esteemed mentor such as Ralph Goodman. We have for the past decade and a half had the pleasure and privilege of conducting research on oceanic bubbles with his help and advice. Over this period, bubble research has been his main focus. It is appropriate that we highlight his leadership and contributions with us on oceanic bubbles. This presentation will provide snapshots of a number of publications and presentations on bubbles since 1993. [Work supported by ONR and NRL.]

Contributed Papers

4:20

2pUW10. Shallow water horizontal signal coherence measurements with use of mobile acoustic sources to create synthetic aperture arrays. Philip Abbot, Ira Dyer, and Chris Emerson (OASIS, Inc., 5 Militia Dr. Lexington, MA 02421)

Horizontal signal coherence length was measured in shallow water using a reciprocal synthetic aperture array created from multiple signals radiated by a mobile acoustic source, and a slowly drifting omnidirectional receiver. The tests were conducted in September, 2008, in shallow water (125 m depth) of the South East China Sea. Source and receiver depths were 61 m. Twelve hyperbolic FM slides per minute were transmitted while the source moved horizontally at 2.5 m/s, maintaining a nearly constant range of ~6 km to the receiver. Received signals were matched-filter processed, phase-corrected to form an equivalent linear array, then coherently combined using a synthetic aperture time delay beamformer. The beamformer output signal gain versus the number of elements in the array, N (using $2 \leqslant N \leqslant 512$ transmissions), was compared to the theoretical limit (20 log, N). The coherence length, defined to be L/λ at a point 3 dB down from the limit, was determined. At 900 Hz, the correlation length was $L/\lambda \approx 40$ (N=7, over 40 s). At 600 Hz, it was $L/\lambda \approx 10$ (N=3 over 8 s). The degradation at 600 Hz is due to substantially smaller signal-to-noise levels at the element level. Partial coherence is observed at both frequencies, even at N=512.

4:35

2pUW11. Attenuation measurements across surface-ship wakes and computed bubble distributions and void fractions: A tribute to the contributions by Ralph Goodman on oceanic bubbles. Steve Stanic (Naval Res. Lab., Stennis Space Ctr., MS 39529, Steve.Stanic@nrlssc.navy.mil), Jerald Caruthers (Univ. of Southern Mississippi, Stennis Space Ctr., MS 39529), Ralph Goodman (Deceased), Edgar Kennedy, and Bob Brown (Naval Res. Lab., Stennis Space Ctr., MS 39529)

A series of three CW-pulsed signals were transmitted across a surface ship wake as the wake aged. Each transmission contained a set of four 0.5-ms-long pulses that ranged over frequencies from 30 to 140 kHz in 10-kHz steps. The acoustic attenuation across wakes that were due to varying bubble-size densities within the wakes were determined experimentally. From those data, estimates of the bubble densities as functions of the speed of the wake-generating ship, the wake's age, and acoustic frequency were calculated. From the bubble-density results, power-law fits and void fractions are calculated. The attenuation measurements were taken at 7.5-m intervals behind the wake-generating ship and continued for about 2 km. The experiment was run for wakes generated at ship speeds of 12 and 15 knot wakes. The bubble densities were observed to have power-law forms with varying parameters with the strongest, for early ages, having an exponent of -3.6 and a void fraction of 4×10^{-7}, and with both diminishing for older wakes, as might be expected. This work was Ralph Goodman's last contribution in underwater acoustics. [Work supported by the Naval Research Laboratory Program Element 62435N and the Office of Naval Research Code 321OA.]

4:50

2pUW12. Observations and theory of dispersion in rip currents at the Scripps Pier Bubble Experiment of 1997. Jerald W. Caruthers (Dept. of Marine Sci. Univ. of Southern Mississippi, 1020 Balch Blvd., Stennis Space Ctr., MS 39529, jerald.caruthers@usm.edu), Steve Stanic (Naval Res. Lab., Stennis Space Ctr., MS 39529), Jorge C. Novarini (Planning System, Inc., Long Beach, MS 39560), and Frank Henyey (Univ. of Washington, Seattle, WA 98105)

Two of the authors (Caruthers and Stanic), with help and advice from Ralph Goodman, were instrumental in planning and conducting the 1997 Scripps Pier Bubble Experiment. Ralph's participation was invaluable in making this effort a success. A number of institutions were involved in the

overall experiments, but the main experiments, conducted by the Naval Research Laboratory, centered on what was termed the "delta frame" [Caruthers *et al.*, J. Acoust. Soc. Am., **106**, 617–625 (1999)]. The delta frame was a triangle 10 m on a side with sources at two apexes and eight hydrophones distributed along the arms. Eight frequencies (39 to 244 kHz) were propagated from the two sources. The NRL personnel and other participants in the experiment published a number of detailed measurements and analyses. In the delta frame experiments, a significant amount of the data was relevant to sound-speed-dispersion analyses. Ralph was always keenly interested in dispersion, but we never fully analyzed, nor published those observations. With that oversight and regret in mind that the relevant analyses were not conducted during Ralph's lifetime, we have revisited that data in his honor. [Work supported by ONR and NRL.]

TUESDAY AFTERNOON, 19 MAY 2009 ALEXANDERS LOUNGE, 2:45 TO 3:45 P.M.

Meeting of Accredited Standards Committee (ASC) S1 Acoustics

P. Battenberg, Chair S1
Quest Technologies, Inc., 1060 Corporate Center Dr., Oconomowoc, WI 53066-4828

R. J. Peppin, Vice Chair S1
Scantek, Inc., 6450 Dobbin Road, #A, Columbia, MD 21045

Accredited Standards Committee S1 on Acoustics. Working group chairs will report on the status of standards currently under development in the areas of physical acoustics, electroacoustics, sonics, ultrasonics, and underwater sound, etc. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAGs for ISO/TC 43 Acoustics and IEC/TC 29 Electroacoustics, take note — those meetings will be held in conjunction with the Standards Plenary meeting at 9:00 a.m. on Tuesday, 19 May 2009.

Scope of S1: Standards, specifications, methods of measurement and test, and terminology in the field of physical acoustics, including architectural acoustics, electroacoustics, sonics, and ultrasonics, and underwater sound, but excluding those aspects of which pertain to biological safety, tolerance and comfort.

TUESDAY AFTERNOON, 19 MAY 2009 DIRECTORS ROOM, 4:00 TO 5:00 P.M.

Meeting of Accredited Standards Committee (ASC) S12 Noise

R. D. Hellweg, Chair S12
Hellweg Acoustics, 13 Pine Tree Rd., Wellesley, MA 02482

W. J. Murphy, Vice Chair S12
NIOSH, 4676 Columbia Pkwy., Cincinnati, OH 45226

Accredited Standards Committee S12 on Noise. Working group chairs will report on the status of noise standards currently under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAGs for ISO/TC 43 Acoustics and IEC/TC 29 Electroacoustics, take note — those meetings will be held in conjunction with the Standards Plenary meeting at 9:00 a.m. on Tuesday, 19 May 2009.

Scope of S12: Standards, specifications, and terminology in the field of acoustical noise pertaining to methods of measurement, evaluation and control, including biological safety, tolerance and comfort, and physical acoustics as related to environmental and occupational noise.

Session 3aAA

Architectural Acoustics: Measurement and Modeling of Scattering Effects

David T. Bradley, Chair
Physics and Astronomy Dept., Vassar College, Poughkeepsie, NY 12604-0745

Invited Papers

8:55

3aAA1. Scattering prediction and measurement in architectural acoustics. David T. Bradley (Phys. and Astron. Depart., Vassar College, 124 Raymond Ave., Poughkeepsie, NY 12604-0745, dabradley@vassar.edu)

Reflected sound energy typically exhibits a combination of both specular and diffuse behavior when acoustic waves encounter a solid surface. For specularly reflected sound, the angle of reflection is equal to the angle of incidence. A diffusely reflected sound is characterized by dispersion in both the spatial and temporal domains. The ability of a surface to diffusely reflect sound is characterized by the amount of scattering in its reflected energy, and can be determined using numerical prediction schemes and experimental measurement techniques. Although the study of scattering surfaces plays an integral role in several fields of acoustics, the development of these measurement and prediction procedures is relatively nascent, particularly in the field of architectural acoustics. An overview of the current state-of-the-art for these procedures will be presented along with an outline of possible future directions in this area of research.

9:15

3aAA2. Measurement and modeling of scattering in underwater acoustics: A potential value to architectural acoustics? Martin Siderius (Dept. of Elect. and Comput. Eng., Portland State Univ., 1900 SW 4th Ave., Portland, OR 97201, siderius@pdx.edu), David Bradley (Vassar College, Poughkeepsie, NY 12604-0745), and Lisa Zurk (Portland State Univ., Portland, OR 97201)

Scattering measurement and modeling techniques in underwater acoustics have undergone significant progress through several decades of research. The knowledge gained from this research can be applied to similar scattering problems found in above water systems, such as those typically seen in architectural acoustics. Underwater environments have inherently complex scattering behavior including multiple propagation paths and boundaries that are rough and spatially variable. Additionally, boundaries are often moving, which can introduce different Doppler shifts on multipath arrivals, referred to as Doppler spread. These complications make modeling underwater propagation and scattering a challenging problem. However, there are currently several models that are routinely used for predicting underwater acoustic behavior. In some cases, deterministic modeling can be used while in others the complexity only allows for a statistical description of the acoustic field. In recent years, acoustic communications have emerged as an important ocean application, and the innovations provided by these systems have pushed the limits of current techniques. An overview of these modeling and measurement techniques will be presented, and the lessons learned from recent advances will be reviewed in the context of identifying commonalities between underwater and architectural acoustics.

9:35

3aAA3. Fractal surfaces: Generation and acoustic scattering prediction. Derek R. Olson and David T. Bradley (Phys. and Astron. Dept., Vassar College, 124 Raymond Ave., Box 2529, Poughkeepsie, NY 12604)

Typical acoustical diffuser design results in diffuse sound scattering only for a limited frequency band, which is problematic given the breadth of the human audible frequency range. Surfaces exhibiting fractal geometries may address this problem due to their self-similarity over multiple scales. Stochastic fractals, such as the random midpoint displacement (RMD) fractal, are well-suited to this usage since they can be mapped onto physical surfaces appropriate for acoustical diffusers. In the current project, virtual RMD fractal surfaces were generated and then constructed using a 3-D printer. A pilot study has been conducted to determine the scattering properties of the fractal surfaces using a numerical prediction scheme carried out using the boundary element method (BEM). Experimental measurements of the scattering properties were also carried out according to ISO 17497. The numerical predictions and experimental measurements were contrasted to improve numerical prediction accuracy and optimize the fractal design parameters of acoustical diffusers. These parameters include surface roughness, fractal dimensionality, number of iterations, and the random number probability distribution function. The comparison of the numerical and experimental results will be presented along with effects of the fractal design parameters.

9:55

3aAA4. A suggested method to be used to measure the scattering coefficient of full-scale samples. Ron Sauro (NWAA Labs., Inc., 25132 Rye Canyon Loop, Santa Clarita, CA 91355, audio_ron@msn.com)

In attempting to follow ISO 17497-1 as a method for measuring full-scale scattering coefficients, it was observed that many of the recommended steps needed to be modified for those full-scale measurements. Variations were tried to aid in these measurements. These variations are described and suggested solutions to the problems observed are presented. It was observed that continuous rotation of the sample, rather than step action rotation, worked better for taking consistent data from the sample. A stable chamber environment is also necessary with humidity reaching at least 50% and temperature variations should not vary by more than 1°C throughout all four parts

of the test. It was also observed that a less directional source than recommended was needed so two dodecahedron loudspeakers were chosen to reproduce a combined weighted noise stimulus source. Six data microphones were placed at random locations and varying heights to collect data. Data were compared between the suggested method and the ISO-17497-1 standard.

10:15—10:30 Break

Contributed Papers

10:30

3aAA5. Investigating multifractals to quantify diffuseness in rooms. Noel W. Hart (Graduate Prog. in Archit. Acoust., Rensselaer Polytech. Inst., 2403 21st St. Apt. 5, Troy, NY, 12180, hartn@rpi.edu) and Ning Xiang (Rensselaer Polytech. Inst., Troy, NY, 12180)

Diffuseness, an important room acoustics parameter, currently has no accurate objective method of quantification. Most methods that claim to be able to quantify it require specialized equipment or guesswork. Multifractals present a recent method that allows diffuseness in rooms to be quantified by analyzing a single room impulse response [Loutridis, JASA **125**(3) (2009)]. This method will be verified by acoustical measurements in rooms of various degrees of diffuseness. Since this method has only been demonstrated in small rooms, its applicability towards larger rooms of different types is tested. Frequency dependency is also unknown and is investigated.

10:45

3aAA6. Prediction and tailoring of steady-state broadband sound fields in enclosures using absorption scaling and energy-intensity boundary elements. Donald Bliss, Krista Michalis, and Linda Franzoni (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708, dbb@duke.edu)

Enclosures with diffuse reflection boundaries are modeled with an energy-intensity boundary element method using uncorrelated broadband directional sources. An absorption-based perturbation analysis shows the spatial variation of acoustic field obeys certain scaling laws. A series expansion in terms of average absorption gives separate boundary integral problems at each order. The lowest-order solution has a uniform level proportional to the reciprocal of the average absorption. The next-order solution is independent of average absorption and primarily responsible for spatial variation of the acoustic field. This solution depends on the spatial distribution of absorption and input power sources, but not their overall level. For the primary spatial variation, the effects of the relative distributions of absorption and input power are linear and uncoupled. These distributions can be expressed in terms of constituent spatial modes corresponding to the ways absorption and input power can be distributed. Solved numerically once for each mode, the acoustic field can be expressed in terms of the modal amplitudes in closed form. These amplitudes can be adjusted to tailor the spatial variation. Examples include how to distribute absorption to minimize sound levels in one location, or how to achieve a uniform interior field. [Sponsor: NSF.]

11:00

3aAA7. Measuring the uniform diffusion coefficient: Synthesized aperture goniometer measurements. Philip W. Robinson, Ning Xiang (Architectural Acoust. Program, Rensselaer Polytechnic Inst.), and Peter D'Antonio (RPG Diffusor Systems, Inc.)

There are currently two methods to measure the diffusion/scattering from acoustical surfaces: reverberation chamber measurement and *in-situ* measurement using a semicircular array of microphones called a goniometer. Each of these methods has advantages and drawbacks. The reverberation chamber method accounts for random incidence, but produces inconsistent results for materials of the same profile but different absorption. It does not distinguish between one-dimensional and two-dimensional surface topologies, since the sample is rotated, and is also inapplicable to high absorption materials. The goniometer measurement quantifies directivity as well as diffusion, but presents challenges related to processing power, as many microphones are required to achieve reasonable resolution. Using a procedure akin to synthesized aperture radar, a 2.5° resolution polar plot of the 180° response is achieved by scanning eight microphones rather than employing a full array of 72. Also, since the direct sound and reflected sound arrive at the receiver in close succession when measuring low incident angles, special processing must be exercised in order to resolve the relationship between the sounds. This research addresses the challenges of the goniometer measurement and attempts to use a synthesized aperture goniometer to derive accurate diffusion and directivity properties of materials.

11:15

3aAA8. Measurement and improvement of the diffuse-reflection coefficients of profiled-wood surfaces. Chris Bibby and Murray Hodgson (Acoust. & Noise Res. Group, SOEH-MECH, Univ. of British Columbia, 3rd Fl., 2206 East Mall, Vancouver, BC, V6T 1Z3 Canada)

An apparatus and procedure were developed for measuring the diffuse-reflection coefficients of surfaces in an anechoic chamber according to the ISO-17497 method. This involves determining the proportion of incoherent energy in the impulse response measured in the presence of the surface. Surface absorption was measured by the reverberation-room method. The sound-absorption and sound-diffusion properties of two existing profiled-wood architectural panels, one with sinusoidal corrugations, the other comprising rectangular sections of different heights, were characterized. Investigations into how to improve the absorption and diffusion characteristics were made, and prototype surfaces tested and evaluated.

11:30

3aAA9. An *in-situ* diffusivity measurement method for diffuser design in concert halls. Jin Yong Jeon and Yong Hee Kim (Dept. of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea)

This study investigates an *in-situ* measurement method to evaluate the diffuseness of a sound field using scale model concert halls. The diffusers were applied to scale model concert halls in which the impulse responses at both the audience and stage area were observed. Both reflection number (RN) and reflection energy (RE) were used as *in-situ* diffusivity indices: RN is defined by the number of reflection rays of which critical amplitude level is within -30 dB after the direct sound, and RE defined by the integrated energy at the normalized impulse responses. Through the scale model measurements, it was observed that half of the side walls near the stage were effective for diffuseness of sounds both in auditorium and on stage. It was also found that the balancing of *in-situ* diffusivity is useful for diffuseness perception, which is subject to the level of the direct sounds.

11:45

3aAA10. Measurement of diffuse sound reflection from an impedance surfaces using one microphone by bayesian inversion. Gavin Steininger (Acoust. and Noise Control Res. Group–Dept. of Mech. Eng., Univ. of British Columbia, Vancouver, BC, Canada) and Murray Hodgson (Univ. of British Columbia, Vancouver, BC, Canada)

This paper discusses the use of inverse methods to find the absorption and diffusion characteristics of surfaces. An impedance surface in an anechoic chamber is excited by a pure tone source above it. The steady-state sound level is measured at n points above the impedance surface. The distribution of the n steady-state sound-pressure levels is assumed to be Gaussian. The set of mean or predicted values for this distribution is generated by finding the modulus of a modified Sommerfeld boundary element solution to the Helmholtz equation. The modification is to add multiple diffusely reflected waves each of which is additionally attenuated by a distribution that is proportional to $\sin(2\theta)\times G(\theta)^{D_\theta}\times H(\phi)^{D_\phi}$, where $G(\theta)$ is the piecewise function $[G(\theta)=\theta/\theta_{Spec},\ \theta\leqslant\theta_{Spec}$, and $[(\pi/2)-\theta]/[(\pi/2)-\theta]_{Spec}$ otherwise] and $H(\phi)=|(1-\phi)/\pi|$. The system of equations is then optimized for the specific impedance of the surface, the normal diffusion coefficient, and the azimuth diffusion coefficient (Z, D_θ, and D_ϕ) using Bayesian

inversion. This process is repeated for two surfaces (painted plywood over 16 inch studs and painted plywood over 16 inch studs with randomly placed wooden blocks) at six frequencies (250, 500, 1000, 2000, 4000, and 8000 Hz).

WEDNESDAY MORNING, 20 MAY 2009 GALLERIA NORTH, 8:00 TO 11.45 A.M.

Session 3aAB

Animal Bioacoustics and Acoustical Oceanography: Autonomous Remote Monitoring Systems for Marine Animals III

Marc O. Lammers, Chair
Hawaii Inst of Marine Biology, Kailua, HI 96734

Invited Papers

8:00

3aAB1. From mysticete song to odontocete echolocation: Monitoring cetacean sounds with high-frequency acoustic recording packages (HARPs). Sean M. Wiggins and John A. Hildebrand (Scripps Inst. of Oceanogr., 9500 Gilman Dr. La Jolla, CA 92093-0205, swiggins@ucsd.edu)

Marine mammals produce a wide range of sounds from 10s of Hz to 100s of kHz. To remotely monitor these various sounds with autonomous systems, instruments with wide bandwidths are required. A general purpose, autonomous, high-frequency acoustic recording package (HARP) is described that can record sounds ranging from low-frequency blue whale (Balaenoptera musculus) song to mid-frequency dolphin whistles and beaked whale sweeps to high-frequency dolphin clicks. In addition to monitoring marine mammals, sounds from fish, contributions from wind and rain, and anthropogenic sources such as ships and sonar also are recorded potentially allowing for the study of cetacean behavioral response to these sounds. HARPs have been deployed for periods of up to one year, in various configurations such as standard moorings, seafloor packages, and arrays, and in remote locations including the Arctic, Bering Sea, Gulf of California, around Hawaii, and offshore of Washington State and southern California. The various configurations of these deployments and the wide range of marine mammal monitoring data they have provided are discussed in addition to lessons learned from these studies. [Work supported by the U.S. Navy CNO-N45.]

8:20

3aAB2. Extending passive acoustic capabilities of autonomous recorders by using multiple hydrophones. Aaron Thode, Robert Glatts (Marine Phys. Lab., Scripps Inst. of Oceanogr., 9500 Gilman Dr., San Diego, CA 92109-0238), and William C. Burgess (Greeneridge Sci. Inc., Goleta, CA 93117)

Low-power data acquisition systems have attained sampling rates large enough to enable multielement hydrophone arrays to be deployed autonomously, at the cost of reducing the individual sampling rate per phone. This paper discusses what theoretical advantages bandlimited multielement recordings can provide over single-hydrophone data, including array gain for increased detection range, interfering noise source rejection, environmental inversion, and biological source tracking. These points are illustrated with data collected from an autonomous four-element vertical array off Queensland, Australia in 2003, and data collected from an autonomous eight-element, 21 m aperture vertical array deployed at 35 m depth in the Beaufort Sea in 2008. These examples also provide insight into the technical and logistical challenges required by such deployments, including deployment and recovery systems that do not endanger or entangle array cables. [Work supported by NPRB, logistics provided by Greeneridge Sciences and Shell Co.]

Contributed Papers

8:40

3aAB3. An autonomous, near-real-time buoy system for automatic detection of North Atlantic right whale calls. Eric Spaulding, Matt Robbins, Tom Calupca, Christopher Clark, Tremblay Tremblay, Amanda Waack, Ann Warde (Bioacoust. Res. Prog., Cornell Univ., 159 Sapsucker Woods Rd., Ithaca, NY 14850), John Kemp, and Kristopher Newhall (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

A moored buoy system for automatic detection of endangered North Atlantic right whale (NARW) upcalls was developed to provide near-real-time information on the presence of whales. The marine components include the WHOI buoy platform (mooring, hydrophone, power system, surface expression, and antennae) and Cornell buoy electronics (housing, analog interface hardware, GPS, embedded computer, detection engine, and telemetry hardware). Shore-side Cornell components include telemetry equipment, server hardware and processing software, database, and interfaces for data annotation, access, and visualization. The buoy hardware/software system is capable of capturing and ranking NARW upcall candidates as 2 s, 2000 Hz sampled audio clips. GPS location, timestamp, and other metadata associated with each audio clip are bundled together and uploaded via satellite for processing. Human analysts regularly annotate incoming data, resulting in a curated database of NARW detections. Periodic "health and status" data allow for confirmation that buoys are functional. Regular voltage reporting helps predict required maintenance. Following initial implementations of the system in 2005, a series of successes and failures have led to system improvements. Deployments have progressed from prototype near-shore units to an operational network continuously monitoring the shipping lanes off Boston to meet ship strike mitigation requirements. Upcoming features and capabilities will be discussed.

8:55

3aAB4. Right whale gunshot calls in the southeastern Bering Sea. Catherine L. Berchok, Laura J. Morse, Phillip J. Clapham (Natl. Marine Mammal Lab., NOAA/AFSC, 7600 Sand Point Way NE, Seattle, WA 98115-6349, Catherine.Berchok@noaa.gov), Holger Klinck, Karolin Klinck, David K. Mellinger (Oregon State Univ. & NOAA Pacific Marine Environ. Lab., Newport, OR 97365), Donald K. Ljungblad (Marine Acoust. Consultants, LLC, Elk Mountain, WY 82324-0006), Sue E. Moore (NOAA/S&T-PMEL, Seattle, WA 98115-6349), Francesco Scattorin (Universita di Pavia, 27100 Pavia, Italy), and Jennifer L. Keating (Zoological Society of San Diego, San Diego, CA 92112-0551)

The eastern population of the North Pacific right whale (Eubalaena japonica) is critically endangered. Following extensive historical whaling in the 19th century and large illegal catches by the USSR in the 1960's, it is likely that this population now numbers in the tens of animals. Little is known about the distribution, movements, migrations, or habitat use of this population, but the existing data suggest that it now occupies a reduced range compared to historical times. In 2007, NMML began conducting a multiyear study of the distribution, abundance, and habitat use of North Pacific right whales in the North Aleutian Basin and southeastern Bering Sea using aerial and vessel surveys. During the 2008 survey, directional sonobuoys were used 24 h a day to record right whale calls and obtain cross-bearings to the whales. Although previous work in the Bering Sea focused on detections of the right whale upsweep call, our findings suggest that the gunshot call is much more ubiquitous. Spatial and temporal trends of gunshot calls will be presented, and changes in gunshot call characteristics relative to observed surfacing sequences will be described. [Work supported by an interagency agreement from the Minerals Management Service.]

Invited Paper

9:10

3aAB5. Environmental acoustic recording system (EARS) buoys for marine animal acoustics. George E. Ioup, Juliette W. Ioup, Arslan M. Tashmukhambetov, Lisa A. Pflug (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, geioup@uno.edu), Joal J. Newcomb (Stennis Space Ctr., MS), Natalia A. Sidorovskaia (Univ. of Louisiana at Lafayette, Lafayette, LA), Stan A. Kuczaj, Grayson H. Rayborn, James M. Stephens (Univ. of Southern Mississippi, Hattiesburg, MS), Christopher O. Tiemann, and Alan Bernstein (Univ. of Texas at Austin, Austin, TX)

EARS buoys were developed as autonomous, moored, underwater recording systems by the Naval Oceanographic Office (NAVOCEANO) to make long-term ocean ambient noise measurements. When the Littoral Acoustic Demonstration Center (LADC) was formed as a consortium of university and U.S. Navy scientists in 2001, the buoys were capable of measuring up to 1000 Hz for 1 yr. LADC added listening to sperm whales to its noise and propagation measurements. NAVOCEANO quickly modified the buoys to measure up to 5859 Hz for 36 days. The buoys, moored at depths from 550 to 950 m in the Gulf of Mexico, produced exceptionally clear recordings of sperm whale echolocation and coda clicks and recordings of other whales. EARS Generation 2 buoys are now capable of recording one channel to 96 kHz, or four channels to 25 kHz, for more than 13 days on four 120 Gbyte notebook disk drives. Experiments in the Gulf of Mexico and the Ligurian Sea have targeted both sperm and beaked whales. Audio results and visualizations of these recordings reveal rich detail of Odontocete clicks and enable new analyses, such as the identification of individual whales from the properties of their clicks. [Research supported by ONR and SPAWAR.]

Contributed Paper

9:30

3aAB6. Localization to verify the identification of individual sperm whales using click properties. Juliette W. Ioup, George E. Ioup, Lisa A. Pflug, Arslan M. Tashmukhambetov (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, jioup@uno.edu), Christopher O. Tiemann, Alan Bernstein (Univ. of Texas at Austin, Austin, TX), Natalia A. Sidorovskaia, Philip Schexnayder (Univ. of Louisiana at Lafayette, Lafayette, LA), Joal J. Newcomb (Stennis Space Ctr., MS), Stan A. Kuczaj, Grayson H. Rayborn (Univ. of Southern Mississippi, Hattiesburg, MS), and Rowena M. Carlson (SPAWAR Systems Ctr. Pacific, San Diego, CA)

Previous research to identify individual sperm and beaked whales from the properties of their echolocation and coda clicks using cluster analysis has been reported. Although reasonably consistent and robust results showing distinct classes (each corresponding to an individual) have been obtained with both self-organizing maps and *K*-means, no independent verification of these identifications has previously been available. However, data from a July 2007 experiment in the Gulf of Mexico should provide enough geometry information to verify some identifications. The Littoral Acoustic Demonstration Center deployed six environmental acoustic recording system buoys, measuring to 96 kHz for 9 days, to record sperm and beaked whale clicks in the northern Gulf of Mexico. Three buoys were in close proximity to enable multisensor detections of single clicks for possible localization, tracking, or bearing estimation sufficient to confirm the identification of individuals from cluster analysis. Displays of cluster classifications versus time coupled with location estimates for the clicks give evidence of the accuracy of the identifications. Though localization serves to verify the classification tool, click identification complements localization by separating sounds from multiple sources when several whales' clicktrains overlap, potentially enabling previously impossible tracking. [Research supported by SPAWAR and ONR.]

Invited Papers

9:45

3aAB7. Detection of killer whale presence using low duty cycle recorders. Jeffrey A. Nystuen (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, nystuen@apl.washington.edu), M. Bradley Hanson, Candice Emmons (NOAA Northwest Fisheries Sci. Ctr., Seattle, WA), Paul Wade (NOAA Alaska Fisheries Sci. Ctr., Seattle, WA), and Jennifer Miksis-Olds (Penn State Univ., State College, PA)

Passive aquatic listeners (PALs) have been developed to make long-term (up to a year) autonomous measurements of the underwater ambient noise in the marine environment. Because of data storage limitations and low-power requirements, PALs have a very low duty cycle, of the order of 1 percent. In fact, they are optimized for acoustic rainfall detection and have been successfully used to measure wind speed and rainfall rate at sea using spectral components of the sound field. These geophysical sounds have relatively long time scales (minutes) compared to vocalizations of marine mammals (seconds) or mooring noise (chain clanking). Data collection was designed to detect and reject short temporal signals. However, these signals include marine mammal vocalizations. Consequently, the sampling strategy was modified to evaluate the temporal content of the data sample and save the temporal data sample only if it contains

a signal consistent with killer whale vocalizations. These sound bites have demonstrated detection of pod-specific vocalizations of resident-type killer whales, transient-type killer whales, bone crunching from likely killer whale predation (eating) on sea lions, and humpback and N. Pacific right whales.

10:05—10:25 Break

10:25

3aAB8. Managing the passive acoustics data deluge. David A. Mann (Coll. of Marine Sci., USF, 140 7th Ave. S., St. Petersburg, FL 33701, dmann@marine.usf.edu)

Passive acoustic recording systems can generate large amounts of data, especially given the increasing availability of inexpensive flash memory. Thus it is easy to collect more data than can be efficiently analyzed. One approach to this problem is to minimize data collection on the recorder by performing real-time automated detection of acoustic targets of interest. An alternate approach, where one does not know what signals may be present, is to record large amounts of data over extended periods of time. DSGlab is an open-source database and data analysis system implemented with MATLAB that is designed to manage large amounts of raw data. The database aspect of the program allows querying of data by latitude, longitude, depth (or altitude). Each recorded data file is tagged with header information including lat/lon/depth, timestamp, sample rate, and calibration. The data analysis section processes the files according to a user-specified signal processing chain, which does not require a knowledge of programming to generate, and returns results to a database and to individual files. The results of signal processing can then be quickly browsed with the data viewer. The system is designed to allow inter-laboratory collaboration and data sharing.

Contributed Papers

10:45

3aAB9. Acoustic rediscovery of right whales in a former whaling area, the Cape Farewell Ground, between Greenland and Iceland. David K. Mellinger, Sharon L. Nieukirk, Karolin Klinck, Holger Klinck, Robert P. Dziak (Cooperative Inst. for Marine Resources Studies, Oregon State Univ. and NOAA Pacific Marine Environ. Lab., 2030 SE Marine Sci. Dr., Newport, OR 97365), and Phillip J. Clapham (NOAA Natl. Marine Mammal Lab., Seattle, WA 98115)

The North Atlantic right whale (*Eubalaena glacialis*) is one of the world's most endangered cetaceans, with only 300–350 animals believed alive. Most right whales are thought to range from Florida to Nova Scotia, though the whereabouts of a significant portion of the population remains unknown in both winter and summer. Here we describe an acoustic survey for right whales near the Cape Farewell Ground, a late 18th-century whaling area. Continuously recording autonomous hydrophone instruments were deployed at five sites for one year in 2007–2008, after which data were analyzed by automatic detection with manual checking for "up" calls of right whales. Over 2000 calls were detected in all, with calls found at all sites but mostly from near the Cape Farewell Ground. The data are consistent with a seasonal migration northeastward toward Iceland in July–August followed by return movement later in the fall, with the last detection in December. At one site, calling occurred in the fall on 17 separate occasions, suggesting a significant number of right whales present. A right whale photo-identified in 2003 at the Cape Farewell Ground was not in the North Atlantic right whale catalog, suggesting the possibility of a previously unidentified stock.

11:00

3aAB10. Sound production and behavior in an eastern North Pacific killer whale population: Implications for autonomous remote monitoring. Marla M. Holt, Dawn P. Noren, Candice K. Emmons (Marine Mammal Program, NOAA/NMFS Northwest Fisheries Sci. Ctr., 2725 Montlake Blvd. East, Seattle, WA 98112, Marla.Holt@noaa.gov), and Anna-Maria Seibert (Univ. of Munich, Planegg-Martinsried, 82152 Germany)

Remote acoustic monitoring is often used to determine the seasonal and spatial distributions of vocal animals, particularly when conditions of other monitoring approaches are limited. Additionally, sound production patterns might be used to infer important activities of free-ranging animals in the absence of other cues. Animals produce sounds during activities such as foraging and breeding, but such patterns are expected to vary between species /populations that are based on a variety of ecological factors. For example, fish-eating killer whales are more vocally active than mammal-eating killer whales, presumably since mammalian prey are sensitive to killer whale sounds. Southern Resident killer whales (SRKWs) are a fish-eating population found in coastal waters of Washington State and British Columbia. The goal of this study was to extend previous work on the behavioral correlates of sound production in SRKWs. We collected behavioral and acoustic data simultaneously over two field seasons and found that click occurrence as well as calling rates (calls/whale/min) were significantly higher when whales were foraging compared to traveling. Such patterns are useful for inferring events from remote acoustic monitoring. These results also indicate that communicative signals are particularly important during foraging activities in this population and have implications for noise disturbance.

11:15

3aAB11. Simultaneous acoustic tag and seafloor acoustic recorder detection of right whale calls in the Bay of Fundy. Susan E. Parks (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804), Christopher W. Clark (Cornell Univ., Ithaca, NY 14850), Mark Johnson, and Peter L. Tyack (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Passive acoustic monitoring is playing a growing role in marine mammal detection. Determining the range of detection for calls of a particular species in a particular location is important to assess the regional coverage provided by individual recording units. This study describes the comparison of right whale calls recorded by digital acoustic recording tags (Dtags) attached with suction cups to North Atlantic right whales and the detection of the same calls using a dispersed seafloor array of autonomous recorders. The seafloor array consisted of 5 units, spaced 6–10 km apart, continuously recording from July 29– August 17, 2005. Dtags were attached to a total of 14 individual right whales during this time period and 7 of these individuals produced a total of 88 tonal calls during tag attachment. The tag and related tracking of the whale provided information on call type, and the timing, depth, and approximate location of the whale producing the call. Tagged whale calls were audible on the seafloor array, and whale-recorder distances provided estimates of the acoustic detection range for right whales in the Bay of Fundy, Canada.

11:30

3aAB12. Annual variability of cetacean occurrence along the northeast coast of Iceland measured using ecological acoustic recorders. E.E. Magnusdottir, M.H. Rasmussen (Husavik Res. Ctr., Univ. of Iceland, Hafnarstett 3, 640 Husavik, Iceland, eem@hi.is), and M.O. Lammers (Hawaii Inst. of Marine Biology, Kailua, HI 96834)

In order to continuously monitor the annual density and habitat use of cetacean species at the NE coast of Iceland, we deployed two seabed mounted ecological acoustic recorders (EARs) in Skjalfandi Bay (66 deg 05 .929 N, 17 deg 25.262 W; 66 deg 03.070 N, 17 deg 39.811 W). The EAR is a microprocessor-based autonomous recorder. Each EAR was programmed to record for 1 min every 15 min. Both EARs will be retrieved every 3–5 months during a 1-year period. A custom Matlab program will be used to automatically detect and analyze target sounds. Target biological sounds re-

corded are primarily those produced by cetaceans and also by pelagic fish. Blue whales (Balenoptera musculus), humpback whales (Megaptera novaeangliae), minke whales (Balaenoptera acutorostrata), white-beaked dolphins (Lagernorhynchus albirostris), and killer whales (Orcinus orca) are well known to visit Skjalfandi Bay. In addition, it has been suggested that cod (Gadus morhua) is spawning in the bay in spring time. The results could reveal annual variability of inter- and intraspecies density of cetaceans and prey within the area. Acoustic monitoring may be an effective means of tracking diurnal and annual activity of cetaceans in a long-term manner and compare the activity to the presence and behavior of prey.

WEDNESDAY MORNING, 20 MAY 2009 BROADWAY I/II, 8:30 TO 11:15 A.M.

Session 3aAO

Acoustical Oceanography and Underwater Acoustics: Environmental Inferences in Inhomogeneous Ocean Environments II

Mohsen Badiey, Chair
College of Marine and Earth Sciences, Univ. of Delaware, Newark, DE 19716-3501

Chair's Introduction—8:30

Contributed Papers

8:35

3aAO1. Information theory application to inversions of acoustic data from a continental shelf environment. David P. Knobles, Robert A. Koch, and Jason D. Sagers (Appl. Res. Labs., The Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713)

This research examines the uncertainty in waveguide parameter values obtained from acoustic inversion of data collected in range-dependent environments. To account for range-dependence in the environmental parameters, the number of inversion parameters must be increased. Generally, from a fixed data sample the uncertainty in the solution for individual parameter values increases as the number of inversion parameters increases, especially if the structure of the propagated acoustic field is sensitive to the additional parameters. This effect can be offset by prior information that reduces the volume of the parameter search space and to a limited extent by larger inversion data samples. How an inversion algorithm accounts for prior information is thus an essential issue to be addressed for range-dependent waveguides. Measurements made on the New Jersey shelf in 70–90 m of water provide an opportunity to study the effects of environmental range dependence on sound propagation and inversion processing. Inversions based on pre-existing geophysical information are combined with a maximum entropy approach to quantify the amount of information on the waveguide parameters that can be extracted from measured propagation over range-dependent tracks. Both towed sources emitting CW lines and impulsive sources are employed. [Work supported by ONR.]

8:50

3aAO2. Probability density functions of modal amplitudes and complex acoustic pressure in fluctuating shallow water waveguides. Kevin D. LePage (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19126 La Spezia, Italy)

The polynomial chaos method is applied to the problem of predicting the probability density functions of complex modal amplitudes and acoustic pressure in the presence of water column sound speed fluctuations in shallow water waveguides. Results for both the intrusive implementation of the polynomial chaos technique, where the governing coupled mode differential equations for the complex modal amplitudes are augmented with the random states of the chaos expansion, and the non-intrusive method, where legacy codes can be run over an ensemble of ocean realizations and the results fitted by a truncated chaos expansion, are shown. Both methods give good agreement with Monte Carlo histograms of the modal amplitudes and the pressure field for slight water column variability, but the non-intrusive formulation shows more robustness for larger variability. The relative merits of PC expansions for the complex modal amplitudes vs the log amplitudes for the complex pressure amplitudes are also discussed.

9:05

3aAO3. Resolving spatial seabed variability by Bayesian inference of seabed reflection inversions. Jan Dettmer, Stan E. Dosso (School of Earth and Ocean Scinces, Univ. of Victoria, Victoria, BC V8W 3P6, Canada), and Charles W. Holland (The Penn State Univ., State College, PA)

This paper considers Bayesian inversion of seabed reflection-coefficient data for multilayer geoacoustic models at several sites with the goal of studying spatial variability of the seabed. 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 Bayesian model selection and comprehensive estimation of data error statistics. Model selection is addressed using the Bayesian information criterion to ensure parsimony of the parametrizations. Data error statistics are quantified by estimating full covariance matrices from data residuals, and *a posteriori* statistical validation is carried out. A Metropolis-Hastings sampling algorithm is used to compute posterior probability densities. Five experimental sites are considered along a track located on the Malta Plateau, Mediterranean Sea, and the inversion results are compared to cores and subbottom profiler sections. Differences between sites that exceed the estimated geoacoustic uncertainties are interpreted as spatial variability of the seabed.

9:20

3aAO4. A simplified model of frequency distortion in ocean acoustic signals by multiple interactions with a moving ocean surface and its use in oceanographic inversions. Stephen D. Lynch and Gerald L. D'Spain (Marine Physical Lab., Scripps Inst. of Oceanogr., 9500 Gilman Dr. 0701, La Jolla, CA 92093-0701)

Received omnidirectional spectra of low frequency narrowband tones measured in a shallow ocean waveguide exhibit Doppler-shifted sidebands at frequencies whose offsets are integer multiples of the dominant surface wave frequencies. These observations suggest that higher order scattering is important. However, higher-order perturbation theory involving only a single surface interaction does not predict the observed relative sideband levels. In particular, the higher order sidebands were often observed to increase in relative level to the main spectral peak during smaller surface wave, and, therefore, smaller kh, conditions. A simplified model is presented that simulates the surface wave-induced frequency distortions as path-length modulation caused by the vertical motions of the surface. The model incor-

porates multiple interactions with the moving surface, and quantitatively predicts the effects on the received spectra of the spatial coherence of the surface waves along the propagation path. The model results are identical to those from higher order perturbation with a single interaction. The simplified nature of this model allows it to be readily used to invert for various properties of the ocean surface wave directional spectrum. [Work supported by ONR, Code 321(US).]

9:35

3aAO5. Particle filter source tracking in a changing geoacoustic environment. Caglar Yardim, Peter Gerstoft, and William S. Hodgkiss, (Marine Phys. Lab., 2038 Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238, cyardim@ucsd.edu)

This paper addresses the problem of tracking the acoustic source parameters such as the depth, range, and speed in evolving geoacoustic environments. It is well known that inaccurate knowledge about the environmental parameters such as the sound speed profile (SSP), water depth, sediment, and bottom parameters may result in significant errors in source parameters. To counter this, a particle filtering (PF) approach is adopted here where the geoacoustic parameters are tracked together with the source location and speed in a range-dependent environment. This allows accurate, real-time updating of the environment the ship is moving in and hence source can be located at any time accurately. As a sequential Monte Carlo technique that can operate on nonlinear systems with non-Gaussian probability densities, the PF is an ideal tracking algorithm to perform tracking of source and environmental parameters and their evolving probability distributions. The algorithm is tested on a sloping environment with the SSP, water depth, and sediment parameters evolving as the ship moves. The change in the water depth created the well-known "source mirage effect," but the PF was still able to track the true source, geoacoustic parameters, and their evolving densities in this spatially varying environment. [Work supported by ONR.]

9:50

3aAO6. Statistics of mode amplitudes in shallow water environments: Effects of random linear internal waves and nonlinear internal wave packets. John Colosi (Naval Postgrad. School, 833 Dyer Rd., Monterey, CA 93943, jacolosi@nps.edu), Andrey Morozov, and Tim Duda (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

In shallow water environments there is presently an incomplete understanding of the relative acoustical importance of the random linear internal-wave field compared to nonlinear internal waves packets. Using the random coupled mode theory of [Creamer (1996)], the nonlinear internal wave mode coupling theory of [Colosi (2007)], and Monte Carlo numerical simulation we are able to demonstrate the important propagation physics when the two internal wave fields act alone, and simultaneously. Importantly, we find that for acoustic frequencies less than roughly 500 Hz and for propagation ranges of order 10s of kms mode propagation though the random internal wave field is very nearly adiabatic, and that cross mode coherences decay rapidly within the first 10 km. A surprising consequence of the decay of cross mode coherence is that the mean intensity will be to first order insensitive to nonlinear internal wave packets if they are located past the decoherence range. It will also be shown that uncorrelated modes have important implications for the approach to saturation. The adiabatic nature of mode propagation though the random internal wave field may have important implications for shallow water acoustic remote sensing.

10:05—10:30 Break

10:30

3aAO7. Inversion for range-dependent water column sound speed profiles in shallow water. 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)

Spatial and temporal variability of the sound speed field in the water column can have a significant impact on acoustic propagation. It is difficult or sometimes impractical to measure fine scale variations in water column properties over an acoustic propagation path. When measurements are not available, water column properties must be approximated. In past work, inverse methods based on acoustic tomography and matched-field processing have been used to estimate mean water column sound speed profiles. The drawback of these methods is their inability to capture fluctuations in the water column sound speed profile. In this work, a perturbative scheme using modal wave numbers is used to obtain range-dependent features of the water column. Based on the work of Rajan *et al.* (1987), a technique has been developed to estimate water column sound speed profiles using historical data to constrain the inversion. This constraint allows for a robust inversion; the result is accurate both when the inputs are noisy and when the bottom is poorly known. The new technique is demonstrated using data from the Shallow Water 2006 (SW06) experiment. Inversion results are compared to *in situ* measurements from the towed CTD chain. [Work supported by NDSEG and ONR.]

10:45

3aAO8. The estimation of geoacoustic parameters via low frequencies (50–75 Hertz) for simulated shallow water test scenarios. A. Tolstoy (ATolstoy Sci., 1538 Hampton Hill Circle, McLean, VA 22101)

This work will demonstrate the success of using only a variety of low frequencies (50–75 Hz) for geoacoustic inversion in simulated SW06 environments. The environments include a variety of scenarios (a thin, medium, or thick sediment layer over a half-space), broadband sources at approximately 1 km range (rge), and a vertical array consisting of 16 phones whose depths are approximately known, i.e., we assume that array element localization was available so that we can estimate phone depths (zphi) to within about 1 m. The source depths (zsou) and water depths (D) are approximately determined via time domain crossings seen at the array [see A. Tolstoy, J. Acoust. Soc. Am. **123**, 3106 (2008)]. Additionally, the ocean sound-speed c(z) is assumed to be range-independent and approximately known, e.g., via measurements or archival data. The use of only the 50–75 Hz portion of the signal results in a need for less than perfect accuracy for zphi and c(z) while still giving excellent accuracy for estimated bottom parameters. Geometric parameters (such as rge, zsou, D, zphi) and c(z) would be more accurately determined by means of higher frequencies but this would require much more CPU time. Uniqueness of the final "solution" is achieved for all but the thin sediment scenario.

11:00

3aAO9. Detection and classification of typhoons using underwater acoustic sensors in the western Pacific Ocean. Barry Ma (Dept. of ECE, Portland State Univ., 1900 SW 4th Ave., Portland, OR 97201, barry@cecs.pdx.edu) and Y. J. Yang (Naval Acad., Tsoying, Kaohsiung, Taiwan)

About two years of underwater ambient noise were collected using Passive Aquatic Listeners (PALs) in the western Pacific Ocean. PALs were deployed using the bottom-mount or surface-mount buoys on the possible typhoon paths. Several episodes of intensive typhoon noise were recorded during the passage of these powerful tropical storms. These acoustic data are compared with other ancillary data for the classification of typhoons. Using the underwater ambient noise to classify the intensity of typhoons is a relatively reliable method. The instrument is placed away from the air-sea interface in a less destructible environment and provides relative *in situ* measurement of the wind speed. The ambient noise recordings from the deployment also include the biology, shipping, and other surface noises, but the noises from the typhoons are the most distinguishable against all other sources of noise. The noise budgets from different deployment sites are also presented.

Session 3aBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Shock Wave Therapy I

Michael R. Bailey, Cochair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105

Thomas J. Matula, Cochair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105-6698

Invited Paper

8:00

3aBB1. Investigation of an ultrasound imaging technique to target kidney stones in lithotripsy. Anup Shah (Dept. of Urology, Univ. of Washington School of Medicine, 1959 NE Pacific St., Box 356510, Seattle, WA 98195, anupshah@u.washington.edu), Marla Paun, John Kucewicz (Univ. of Washington, Seattle, WA 98105), Oleg A. Sapozhnikov (Moscow State Univ., Moscow 119992, Russia), Manjiri Dighe (Univ. of Washington School of Medicine, Seattle, WA 98195), Hunter A. McKay (The Polyclinic, Seattle, WA 98122), Mathew D. Sorensen (Univ. of Washington School of Medicine, Seattle, WA 98195), and Michael R. Bailey (Univ. of Washington, Seattle, WA 98105)

Localization of kidney stones and targeting for lithotripsy can be challenges especially with ultrasound. However, twinkling artifact has been observed where Doppler ultrasound imagers assign color to the stone. We report a preliminary investigation from our observations in a porcine model and hypothesize why this artifact occurs. Glass beads, cement stones, and human stones were surgically placed into the renal collecting system through the ureter. The stones were imaged using several transducers and ultrasound imagers. In all cases, the twinkling artifact of the stone was observed, and its appearance and radiofrequency signature were unique from those of blood flow. Calcium oxalate monohydrate stones and smooth stones were not more difficult to image, contrary to previous reports. Increasing gain or placing the focal depth distal to the stone enhanced the artifact, but other user controls showed little effect. Twinkling started at the lateral edges of the stone and spread over the stone as gain was increased. The evidence supports the hypothesis that small motions induced by radiation force or elastic waves in the stone cause changes in received backscatter, particularly at imaging angles oblique to the stone surface. [Work supported by NIH DK43881 and NSBRI SMST01601.]

Contributed Paper

8:15

3aBB2. Shock wave lithotripsy treatment at 60 shock waves per minute is an appropriate protocol to minimize tissue injury. Bret Connors, Andrew Evan, Philip Blomgren, Rajash Handa, Cynthia Johnson, James McAteer (Dept. of Anatomy, IU School of Medicine, Ms-5055, 635 Barnhill Dr., Indianapolis, IN 46202), Lynn Willis (IU School of Medicine, Indianapolis, IN 46202), and James Lingeman (Int. Kid. Stone Inst., Indianapolis, IN 46202)

Previous work has shown that the size of the renal hemorrhagic lesion caused by extracorporeal shock wave lithotripsy (SWL) is greatly reduced when shock waves (SWs) are delivered at 30 SWs/min instead of at the more commonly used rate of 120 SWs/min ($0.08 \pm 0.02\%$ versus $3.93 \pm 1.29\%$ of functional renal volume, $P=0.006$). However, slowing the SW rate to 30 SWs/min greatly increases treatment time. Consequently, we investigated whether an intermediate rate of 60 SWs/min also reduces lesion size compared to treatment at 120 SWs/min. Methods: one kidney in 12 juvenile pigs was treated at 120 SWs/min ($n=7$) or at 60 SWs/min ($n=5$) with an unmodified Dornier HM-3 lithotripter (2000 SWs, 24 kV). After lithotripsy, the treated kidneys were fixed and removed to determine lesion size. Results: SWL at 60 SWs/min significantly reduced the size of the morphological lesion compared to 120 SWs/min ($0.42 \pm 0.23\%$ versus $3.93 \pm 1.29\%$, $P=0.034$). This reduction in lesion size was comparable to the lesion produced at 30 SWs/min. Conclusions: SWL applied at 60 SWs/min produces significantly less renal injury than SWL at 120 SWs/min, and produces an equivalent level of injury as that seen at 30 SWs/min.

Invited Papers

8:30

3aBB3. Renal heme oxygenase-1 upregulation after shock wave lithotripsy. Daniel L. Clark, Bret A. Connors, Andy P. Evan, Rajash K. Handa, and Cynthia D. Johnson (Dept. of Anatomy Cell Biology, Indiana Univ. Sch. of Medicine, 635 Barnhill Dr., Indianapolis, IN 46202, danlclar@iupui.edu)

Shock wave lithotripsy (SWL) is known to induce oxidative stress and a rapid inflammatory response in renal tissue. The purpose of this study was to determine the location and quantify markers for oxidative stress and inflammation in a porcine model of acute SWL-induced renal injury. The lower pole calyx of left kidneys of female pigs received 2000 shock waves at 24 kV/2 Hz from a Dornier HM3 lithotripter and subsequently monitored for 4 h. Heme oxygenase-1 protein (HO-1) was measured in renal microsomes by Western blot. Renal medulla HO-1/β-actin ratios were 1.186 ± 0.48 for treated pole (F2), 0.293 ± 0.23 for treated kidney upper pole, and 0.122 ± 0.075 for contralateral kidney ($n=6$). Thus, we observed an eightfold induction of HO-1 in renal medulla at F2 compared to the upper pole medulla of the treated kidney and the contralateral kidney medulla. In contrast, renal cortex at F2 did not show a similar HO-1 induction. The highly localized HO-1 induction in SWL-treated renal medulla parallels our previously reported finding of interleukin-6. These data together suggest that an early inflammatory response and oxidative stress appear primarily in the treated renal medulla within 4 h after SWL. [Work supported by NIH].

8:45

3aBB4. Low-energy shock wave pretreatment results in greater vasoconstriction and less injury in the kidney compared to high-energy shock wave lithotripsy treatment alone. Rajash K. Handa, Andrew P. Evan, Bret A. Connors (Dept. of Anatomy and Cell Biology, Indiana Univ. Sch. Med., Indianapolis, IN 46202), Lynn R. Willis, Sujuan Gao (Indiana Univ. Sch. Med., Indianapolis, IN 46202), Marla Paun, and Michael R. Bailey (Univ. of Washington, Seattle, WA 98115)

Pretreating the kidney with low-energy shock waves (SWs) before administering a clinical dose of high-energy SWs has been found to greatly decrease the hemorrhagic lesion that normally results from high-energy shock wave lithotripsy (SWL) treatment alone. We tested the hypothesis that this novel pretreatment SWL protocol results in a greater and/or earlier renal vasoconstriction than clinical SWL alone, which could potentially explain the reduced bleeding in the kidney. Anesthetized pigs were treated with a clinical dose of SWs (2000 SWs, 24 kV, 120 SWs/min) or the pretreatment protocol (500 SWs, 12 kV, 120 SWs/min + 3-min pause + 2000 SWs, 24 kV, 120 SWs/min) using the HM3 lithotripter. Sonography (color and pulsed ultrasound) was used to locate a resistance artery within the SW-treated pole of the kidney and to take resistance index (RI) measurements from the blood vessel to estimate the resistance /impedance to blood flow. The results demonstrated that RI increased (renal vasoconstriction) earlier and greater during the pretreatment protocol compared to normal clinical SWL. Constricted blood vessels would likely be less prone to rupture by SWs and, if broken, would bleed less—resulting in a smaller hemorrhagic lesion.

Contributed Paper

9:00

3aBB5. Efficiency of spark discharge in electrohydraulic lithotripsy. Yuri A. Pishchalnikov, Mark M. Kaehr, R. Jason VonDerHaar, Irina V. Pishchalnikova, and James A. McAteer (Dept. of Anatomy and Cell Biology, Indiana Univ. Sch. of Medicine, 635 Barnhill Dr., Indianapolis, IN 46202, yura@anatomy.iupui.edu)

Electrohydraulic lithotripters and SWT devices generate shock waves by discharge of a high-voltage capacitor through submerged electrodes. As the electrodes age, the interelectrode gap widens. How this affects the efficiency of spark generation was studied using a research HM3-clone lithotripter. Widening of the interelectrode gap (~0.3 mm with new electrodes; ~2.5 mm after 4000 discharges) increased the lag-time to breakdown (~0 to ~30 μs, respectively). Increased lag-time as electrodes aged was associated with partial discharge of the capacitor (leakage promoted by ~0.6 mS conductivity of the surrounding water), such that the average energy remaining at the capacitor at the moment of breakdown was reduced four times compared to new electrodes. However, with new electrodes almost 90% of the energy was lost in the circuitry rather than in the spark, as the resistance of the spark (R~0.03 Ω) was much smaller than the resistance of the remainder of the circuit—including the high-voltage switch and connective wires (r~0.23 Ω). With old electrodes, most of the energy was released in the spark, as the electric efficiency $R/(R+r)$ approached ~90%. The product of electric efficiency and capacitor energy agreed closely with both the acoustic energy and the volume of the primary bubble produced at spark discharge. [Work supported by NIH-DK43881.]

Invited Paper

9:15

3aBB6. Bubble dynamics with tissue confinement in shock-wave lithotripsy. Jonathan B. Freund (Mech. Sci. Eng., Univ. of Illinois at Urbana–Champaign, 1260 W Green St., MC-244, Urbana, IL 61801, jbfreund@illinois.edu) and Ratnesh Shukla (Univ. of Illinois at Urbana–Champaign, Urbana, IL 61801)

Estimates are made of the effect of confinement by tissues on the action of small bubbles when subjected to strong pressure waves. The applications of interest are biomedical procedures involving short strong ultrasound bursts or weak shocks of the kind delivered in shock-wave lithotripsy. Confinement is anticipated to be important in suppressing mechanical injury and slowing the rate of its spread. We consider bubbles in a liquid such as blood within a small vessel in the tissue. A generalization of the Rayleigh–Plesset equation allows us to estimate the effect of the elasticity and viscosity of the surrounding tissue. Ranges of soft-tissue properties are estimated from a variety of different measurements available in literature. Solutions suggest that elasticity is insufficient to significantly alter bubble dynamics, but that viscosities from the mid-to-high range of those suggested might play a significant role in suppressing bubble action. Simulations in two space dimensions of a shocked bubble in a waterlike fluid interacting with a viscous material show that the much more complicated bubble jetting dynamics in this configuration is also significantly suppressed. The dynamics of this suppression is investigated.

Contributed Paper

9:30

3aBB7. Simulation of three–dimensional crack profiles in shock wave lithotripsy. Maddegedara Lalith Lakshman Wijerathne, Hide Sakaguchi (JAMSTEC, Yokohama Inst. for Earth Sci., 3173-25 Showa-machi, Yokohama 236-0001, Japan.), Kenji Oguni, and Muneo Hori (Univ. of Tokyo, Bunkyo-Ku, Tokyo 113-0032, Japan.)

Some of the experiments reported by Xufeng Xi and Pei Zhong in 2001 were simulated to explore the sources of crack initiation and propagation in shock wave lithotripsy (SWL). An interesting result they reported was the T-shaped crack pattern in plaster of paris cylinders, subjected to transverse pressure pulses. This nontrivial crack pattern and some others were simulated, in 3-D, using a numerical technique called PDS-FEM. According to this 3-D simulations, it was found that the interference of reflected and focused high pressure pulse, by the distal surface, and the lagging low tensile phase created high tensile regions in the cylindrical sample. These high stress regions initiate and drive two horizontal and vertical cracks leading to the T-shaped crack. Due to the lack of data, a typical lithotripter pressure wave profile and material properties for plaster of paris were used. The observed crack patterns were successfully regenerated in 3-D, irrespective of these approximations. The distances between the vertical crack and the distal face for the crack patterns of the experiments and numerical simulations are nearly $0.34\times d$ and $0.35\times d$, respectively (d=diameter). This quantitative comparison shows the potential applications of PDS-FEM for exploring and further development of SWL technology.

Invited Papers

9:45

3aBB8. Shock wave interaction with cavitation bubble clusters generated in lithotripsy. Derek C. Thomas, Todd A. Hay, and Mark F. Hamilton (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713-8029)

Collapse of cavitation bubbles generated in shock wave lithotripsy is essential to the comminution of kidney stones. After passage of the shock, however, if insufficient time has elapsed for the bubbles to dissolve, acoustic scattering from residual bubbles shields the stone from the next shock. Therefore, cavitation also limits the maximum firing rate. As important as collective bubble dynamics are to the efficacy of shock wave lithotripsy, behavior of the bubble clusters is among the least understood physical processes involved. A model was developed previously for the pulsation and translation of dynamically coupled spherical gas bubbles [Ilinskii *et al.*, J. Acoust. Soc. Am. **121**, 786 (2007)]. This model was augmented to account for interaction with the focused shock wave incident on the cluster at the order of the Born approximation. Diffraction of the shock wave is described by the KZK equation. The model predicts translation and coalescence of the bubbles due to transmission of the shock through the cluster. Predicted also is the acoustic wave scattered from the cluster, which in the forward direction is the leading order effect of the cluster on the transmitted shock wave. [Work supported by the ARL McKinney Fellowship in Acoustics and NIH DK070618.]

10:00—10:15 Break

10:15

3aBB9. Shock wave lithotripsy can alter urinary acid-base pH regulation. Rajash K. Handa, Bret A. Connors, Cynthia D. Johnson, Andrew P. Evan (Dept. of Anatomy and Cell Biology, Indiana Univ. Sch. Med., Indianapolis, IN 46202), Mahesh C. Goel, James E. Lingeman (Methodist Hospital Inst. for Kidney Stone Disease, Indianapolis, IN 46202), Elaine M. Worcester, and Fredric L. Coe (Univ. of Chicago, Chicago, IL 60637)

The renal papilla is particularly susceptible to injury by shock waves (SWs). Since the papilla is vital for urinary pH regulation, we hypothesized that shock wave lithotripsy (SWL)-induced damage of papillae will alter the pH of the urine. Experiments were conducted in four female adult Gottingen minipigs. The HM3 lithotripter delivered a total of 8000 SWs (24 kV, 120 SWs/min) among all papillae of one kidney, while the untreated, opposite kidney was used as control. The pigs were allowed to recover and 5 weeks later they were anesthetized. Serial urine collections were obtained from the SW-treated and untreated kidney. Urine was collected under mineral oil and immediately tested for its pH. Both kidneys were then perfusion fixed for histological analysis. Urinary pH was 0.17 units greater ($P<0.05$) in the SW-treated kidney compared to the untreated kidney. Histology showed scarring in all regions of the SW-treated kidney, with thick ascending limbs and collecting ducts having grossly abnormal morphologies indicative of cellular proliferation. These results suggest that SWL-induced injury can cause remodeling of nephron structures and long-term impairment in the regulation of urinary pH.

10:30

3aBB10. Ureteroscopic ultrasound technology to size kidney stone fragments: Proof of principal using a miniaturized probe in a porcine model. Mathew D. Sorensen, Anup Shah (Dept. of Urology, Univ. of Washington Sch. of Medicine, 1959 NE Pacific St., Box 356510, Seattle, WA 98195, mathews@u.washington.edu), Michael S. Canney (Univ. of Washington, Seattle, WA 98105), Oleg A. Sapozhnikov (Moscow State Univ., Moscow 119992, Russia), Joel M. H. Teichman (Univ. of British Columbia, Vancouver, BC V6Z 1Y6, Canada), and Michael R. Bailey (Univ. of Washington, Seattle, WA 98105)

Background: the ability to measure stone fragment size could help prevent attempting to extract too large a stone fragment. We evaluated the ability of a 1.2 mm (3.6 French) ultrasound probe to measure stone fragments in a porcine kidney. Methods: 15 human stones of three types (five each calcium oxalate, cystine, calcium phosphate) sized 3–7 mm were placed deep in a porcine kidney collecting system. The sound speed of each stone type was determined using a separate reference stone. A 2 MHz, 1.2 mm needle hydrophone was used to send and receive ultrasound pulses. Stone thickness was calculated as $d=c*t/2$ by determining the signal transit time through the stone, t, and the stone sound speed, c. Calculated stone thicknesses were compared to digital caliper measurements. Results: Stone size was determined for all 15 stones. Correlation between ultrasound-determined thickness and caliper measurements was excellent ($r^2=0.90$, $p<0.0001$) with ultrasound performing well in all three stone types. All stone measurements were accurate within 1 mm, and ten (66%) stone measurements were accurate within 0.5 mm. Conclusions: Ultrasound-based measurements are accurate and precise using a 3.6 French probe with stone fragments placed deep in a porcine kidney. [Work supported by Grants NIH DK43881 and NSBRI SMST01601.]

Contributed Papers

10:45

3aBB11. Ellipsoidal-reflector high intensity focused ultrasound system with integrated safety monitoring. Joshua Samuels and Vesna Zderic (Dept. of Elec. and Comp. Eng., The George Washington Univ., 801 22nd St. NW, Washington, DC 20052, zderic@gwu.edu)

High intensity focused ultrasound (HIFU) transducers can be created using concave piezo elements, acoustic lenses, phased arrays, or ellipsoidal reflectors. By using a 2 MHz cylindrical piezoelectric crystal and a brass ellipsoidal reflector, it was possible to use the base of the transducer to house specific devices which can be integrated into a safety system for the device. This system assists in focal point positioning, ablation confirmation, as well as assuring proper flow of coolant through the attached water pillow to couple the device. These methods combine to protect the patient and the device itself from improper use or dangerous positioning of the focal point within the tissue. The proposed system utilizes an unfocused 3.5 MHz ultrasound transducer to act as a rangefinder, determining the distance of the tissue from the transducer. A simple graphical user interface allows the maximum allowed depth of ablation to be user defined and a warning displayed if needed. A water pillow with integrated flow meter is used to provide cooling and coupling to the tissue. There is also a housed video camera in the device which could provide visual confirmation of ablation if used laparoscopically or for superficial lesions.

11:00

3aBB12. *In vivo* simulation of shock wave lithotripsy: Wave focusing in inhomogeneous materials. Jeffrey Krimmel and Tim Colonius (Caltech 1200 E California Blvd., MC 104-44, Pasadena, CA 91125, jkrimmel@caltech.edu)

Shock wave generation and focusing in electrohydraulic, electromagnetic, and piezoelectric lithotripters are simulated numerically. Inhomogeneities associated with wave propagation *in vivo* are modeled using data from The Visible Human Project. The approximate time-domain relaxation model of Yang and Cleveland (2005) was used to simulate attenuation and dispersion in the tissuelike media. The simulations utilize a MUSCL-type shock capturing scheme with adaptive mesh refinement (AMR). *In vitro* focal region pressure measurements compare favorably with available experimental data for each lithotripter type. Preliminary *in vivo* simulation results show significant dispersion associated with the nonuniform sound speed, including increases of up to 50% in focal pressure amplitudes and the presence of multiple pressure minima/maxima in the focal waveform. We will also discuss the extension of the AMR framework to account for cavitation via a previously developed ensemble-averaged continuous two-phase flow model.

11:15

3aBB13. Kidney stone tracking *in vitro* using an acoustic triangulation paradigm. Jonathan M. Kracht and Robin O. Cleveland (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

During shock wave lithotripsy, stones undergo motion which can place them outside the focal zone of the lithotripter. This results in shock waves being delivered that do not impact the stone but may injure tissue. Tracking stones using diagnostic ultrasound imaging (~4 MHz) has proven to be challenging. Here we employed an array of seven relatively low-frequency (~600 kHz) elements to detect scattered signals from an artificial kidney stone in the presence of a tissue phantom. Using an optimization routine, the time of flight to each element and array geometry were used to determine the most likely location of the stone. A comparison of threshold crossing and cross-correlation for detecting signal arrival indicated that while the former yielded faster computation time, the latter was more robust to noise. Stone position was determined to within 2 mm for locations within 10 mm of the focus. For distances beyond 10 mm, the optimization routine was not able to reliably predict stone location but could indicate that the stone was not within the focal region. [Work supported in part by NIH.]

11:30

3aBB14. Lung hemorrhage produced by exposure to underwater acoustic impulses. Diane Dalecki, Sally Z. Child, and Carol H. Raeman (Dept. of Biomedical Eng., and the Rochester Ctr. for Biomedical Ultrasound, 310 Goergen Hall, Univ. of Rochester, Rochester, NY 14627, dalecki@bme.rochester.edu)

Underwater acoustic impulse sources are used in various commercial and military applications. The objective of this study was to investigate the response of mammalian lung to exposure to underwater acoustic impulses. A laboratory-based acoustic exposure system was developed to investigate the bioeffects of underwater impulses. Underwater acoustic impulses were generated using a ten cubic inch air gun operated in a large water tank. Underwater acoustic impulses with peak pressure amplitudes ranging from 0 (sham) to 110 kPa were investigated. Experiments were performed with anesthetized adult mice. Each exposure consisted of five acoustic impulses with approximately 20 s between impulses. Mammalian lung hemorrhage can occur in response to exposure to underwater acoustic impulses. The pressure threshold for lung hemorrhage resulting from five air gun impulses was approximately 50 kPa. The extent of lung damage increased with increasing pressure amplitude. The results of this work are relevant to establishing safety guidelines for swimmers and divers exposed to underwater sound fields.

WEDNESDAY MORNING, 20 MAY 2009 COUNCIL SUITE, 9:00 TO 11:15 A.M.

Session 3aEA

Engineering Acoustics, Structural Acoustics and Vibration, Noise, and Acoustical Oceanography: Acoustical Engineering of Wind Turbines

Thomas R. Howarth, Chair
Naval Sea Systems Command Division Newport, Newport, RI 02841

Chair's Introduction—9:00

Invited Papers

9:05

3aEA1. Wind turbines: Why they are noisy and what to do about it. Frits van den Berg (GGD, P.O. Box 2200, 1000CE Amsterdam, The Netherlands, fvdberg@ggd.amsterdam.nl)

Sound from modern, tall wind turbines is related to sleep disturbance and is, per decibel, more annoying than sound from common sources such as road or air traffic. Surveys among residents indicate that there are several reasons for this: the swishing character of the sound, the intrusiveness at night, and the visibility of the wind turbine(s). Residential reaction is also determined by the attitude towards wind turbines in the landscape and by economical benefits. From acoustical research different explanations have arisen as to why the sound is amplitude modulated. For a distant observer it is the result of the change in trailing sound level due to the change in wind that the revolving blades encounter. High night-time sound levels are due to an increase in rotor height wind speed simultaneously with a decrease in near-ground wind speed. Wind turbine design has been directed to higher yields per turbine. Less attention has been given to low noise blades and rotor speed reduction. Low noise design could also include a reduction in the modulation amplitude and more sophisticated speed reduction; viz., at high annoyance conditions. Also, involving residents in wind farm planning may have a high potential to reduce annoyance.

9:30

3aEA2. An overview of the acoustical activities at the National Wind Technology Center. Arlinda Huskey (NREL, NWTC, 1617 Cole Blvd., MS 3911, Golden, CO 80401, Arlinda_Huskey@nrel.gov), Moriarty Patrick, Jeroen van Dam (NREL, NWTC, Golden, CO 80401), Eric Simley, and Scott Palo (Univ. of Colorado)

The National Wind Technology Center (NWTC) has been involved in wind turbine acoustics since the 1980s. The areas of work include standards development, measurement and analysis, prediction code development, and acoustic arrays. Most measurements are conducted to the IEC 61400-11, a standard specifically for acoustic measurements of wind turbines. The NWTC is involved in the development of this standard. The standard results in reports of overall levels as well as tonality. The NWTC has also developed a semiempirical acoustic noise prediction code for wind turbines. Validation of this code with test data (both wind tunnel and field measurements) shows reasonable prediction accuracy for both airfoil self noise and turbulent inflow noise. Lastly, to assist with wind turbine noise reduction and to identify the loudest components an acoustic array was developed to better locate noise on wind turbines. This paper will describe the NWTC's past and current activities in wind turbine acoustics including test methods and results.

9:55

3aEA3. Ocean acoustic noise budgets: Application to the environmental assessment of offshore wind power generation. James H. Miller, Gopu R. Potty, Andres Nunez Perez (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882, miller@uri.edu), Kathleen Vigness Raposa (Marine Acoust., Inc., Middletown, RI 02842), and Jeffrey A. Nystuen (Univ. of Washington, Seattle, WA 98105-6698)

A noise budget is a listing of the various sources of acoustic noise and their associated ranking by importance. A number of different types of budgets can be conceived using various acoustic measures such as intensity, energy, or duration of maximum amplitude level. These budgets are typically parameterized by frequency and are usually computed over 1/3 octave bands. As part of the environmental assessment of the proposed offshore wind power generation project under the Rhode Island Special Area Management Plan (SAMP), noise measurements were made using the Passive Acoustic Listener (PAL) systems off the coast of Rhode Island prior to the installation of any wind power facilities. Two PALs were deployed within two miles of Block Island in water depths of 20 m from October 6 to November 11, 2008. The data included noise spectra and source identification every 3 min. Short snapshots of unusual sounds were also recorded. From this data, the ocean acoustic noise budget is computed with contributions from shipping, wind/waves, marine mammals, and rain from 500 Hz to 50 kHz. The ship noise data is correlated with ship traffic data from the Automatic Identification System (AIS). [Funding provided by the Rhode Island Office of Energy Resources.]

10:20

3aEA4. Prediction methodologies for tonal and broadband noise from horizontal-axis wind turbines. Brent Paul, Rudolph Martinez, Donald Thompson, and James Uhlman (CAA-Alion Corp., 84 Sherman St., Cambridge, MA 02140 bpaul@alionscience.com)

This paper describes a set of computational-analytic predictive tools for the following mechanisms of aerodynamic sound and vibration for a wind turbine's blades in ducted or open arrangements: (a) Infrasound and tonal low-frequency nearfield noise from the interaction of blades with a nonuniform meanflow that is steady in its own reference frame; (b) mid-frequency broadband noise from the "haystacked" chopping of an incident turbulent freestream; (c) high-frequency noise from self-generated turbulence near the trailing edge of each blade's suction side. Contributors to tonal mechanism (a) include the rotor's static loads as required by power generation (nearfield "Gutin" component of infrasound), and operation in a local wind shear or in the mean wake of the turbine's tower when set up as a downstream machine. These predictive tools include models of the spatially nonuniform mean and turbulent flows that typically strike a wind turbine's blades, as well as preprocessors of that information when made available by either field measurements or off-line calculations. The codes' physical content includes the possibility of uneven spacing for the turbine's blades and for their interaction with statistically inhomogeneous turbulence for mechanism (b). Sample predictions compare the codes' estimates of sound production to measurements from the open literature.

Contributed Papers

10:45

3aEA5. Methodologies to address limitations of the ISO9613-2 standard for wind turbine noise propagation modeling. Erik J. Kalapinski (Tetra Tech EC Inc., 133 Federal St., Boston, MA 02110, erik.kalapinski@tteci.com)

The accuracy of the noise propagation modeling results is dependent on: (1) the sound source data entered into the model as reported by wind turbine equipment manufacturers as specified under per IEC 64100-11 standard; (2) site specific topographical, meteorological, and terrain conditions; and (3) the propagation algorithms of the acoustic modeling standard that are applied. For wind energy facilities sited in the US, the ISO 9613-2 standard is most commonly used due to its proven effectiveness to accurately predict the sound levels at a distance for industrial and commercial projects in an outdoor environment. The propagation algorithms employed by the ISO 9613-2 standard account for geometrical divergence, atmospheric absorption, ground attenuation, screening effects, and favorable sound propagation conditions that would occur during omnidirectional downwind meteorological conditions. The purpose of this paper is to identify limitations inherent in the ISO 9613-2 standard to the specialized application of wind turbine acoustics and nonstandard atmospheric conditions. Methodologies are presented to more accurately describe atmospheric and lateral attenuation effects, which can result in significant uncertainties in the noise levels calculated both in proximity and at distant receptor locations.

11:00

3aEA6. Vibration analysis of turbine low-pressure bladed-disk in 500-MW fossil power plant. Dooyoung Lee, Youngchae Bae, Heesoo Kim, Yookryun Lee, Hongil Choi, and Hyun Lee (Hydro & Fossil Power Generation Lab., KEPRI, 103-16 Munji Yusung Daejon, Korea, 305-380, dylee@kepri.re.kr)

Many failures of turbine blades are caused by the coupling of aerodynamic forcing with bladed-disk vibration characteristics. Vibration analysis for an individual turbine blade does not include all the coupled characteristics of whole bladed-disk system. This paper shows the coupled vibration characteristics of a low-pressure (LP) turbine bladed-disk in a 500-MW fossil power plant for the purpose of confirming equipment integrity. In order to verify the finite element model, analytic and experimental modal analyses for a single blade were performed and reviewed with boundary conditions not only where both ends were free to move but also where its root was

fixed. Natural frequencies and corresponding mode shapes were calculated for the bladed-disk first, and then proximity of modes to sources of excitation was assessed by means of interference diagrams to examine resonance. The result of analysis was ascertained by being compared with a field test for the bladed-disk without centrifugal and aerodynamic forces. In addition, dynamic stress as well as steady stress which results from centrifugal and steady-state steam forces was also calculated at dangerous modes of operation. It is expected that these dynamic characteristics will be used effectively to identify the root causes of blade failures and to perform prompt maintenance.

Session 3aED

Education in Acoustics and ASA Student Council: Project Listen Up

James M. Sabatier, Chair
National Ctr. for Physical Acoustics, Univ. of Mississippi, University, MS 38677

Chair's Introduction—8:25

Invited Papers

8:30

3aED1. A simple experiment for understanding resonant air columns. Theodore F. Argo, IV, Chad A. Greene, and Preston S. Wilson (Mech. Eng. Dept. and Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78712-0292, targo@mail.utexas.edu)

One goal of Project Listen Up is to enhance children's understanding of acoustics through engagement in basic acoustics experiments. This work proposes an experiment that examines the resonant frequencies of an air column. Equipment for this experiment consists of a small tunable oscillator circuit connected to a loudspeaker, which can also be used for other experiments in the kit. The experimenter will provide a cardboard tube, rubber band, plastic wrap, and sand. Resonance is demonstrated by placing the loudspeaker at the base of the tube, covering the top of the tube with a plastic wrap membrane, and placing sand grains on top of the membrane. As frequency is adjusted, the sand grains are excited at the resonance frequencies of the tube. Frequencies at which sand motion is observed can be compared to a simple algebraic model, which relates the length of the tube to the resonance frequencies of the system. An optional advanced experimental track investigates the effect of varying the length of the tube. This experiment develops intuition for the relationship between frequency and wavelength and builds understanding of resonant systems.

8:45

3aED2. Sound speed measurements in air using a variable sound source and tubes. Stanley A. Cheyne and Walter C. McDermott (Dept. of Phys. and Astronomy, Hampden-Sydney Coll., Hampden-Sydney, VA 23943)

Sound speed measurement using standing waves and tubes is a standard experiment in every introductory physics laboratory. It is an experiment that is relatively simple to understand both conceptually and mathematically. The idea behind this work is to reproduce the same experiment, at low cost, making it suitable for project listen up. Two approaches were made. The first was to design and construct a small, low cost, variable sound source that could be used to determine the resonances of several fixed length tubes. The device is simple in design and easy to use. As the frequency is varied, the resonances can be determined by listening with the ear. Once done, the sound speed can be easily calculated. The second method was to use a free, online, function generator as the sound source and repeat the experiment as previously discussed. Results of both methods will be given and demonstrated along with pros and cons of each.

9:00

3aED3. Why we have two ears—a hands-on experiment comparing monaural and binaural hearing. Ralph T. Muehleisen (Civil, Environ., and Architectural Eng., Illinois Inst. of Technol., 3201 S. Dearborn St., Rm. 228, 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. This paper describes a simple experiment in which two individuals work together to investigate the sound localizing abilities of a human using monaural and binaural hearing. In the experiment, one person sits blindfolded in the center of a circle and acts as a listener. The other person moves a sound source around a circle, equidistant from the listener while the listener attempts to locate the direction of the source. The experiment is repeated for both monaural (with an earplug in one ear) and binaural (using both ears) hearing and for impulsive and low frequency tonal sources. The experiment is best attempted in an environment with very few nearby reflecting sources such as the outdoors or a very large room.

9:15

3aED4. Two-dimensional normal modes: Experiment and finite element analysis. Uwe J. Hansen (Dept. of Chem. & Phys., Indiana State Univ., Terre Haute, IN 47809)

Bending wave propagation on a two-dimensional plate is reduced to standing waves with the imposition of boundary conditions. These standing wave patterns can be observed with sand, which accumulates near the nodal lines. These sand patterns are usually called Chladni patterns after Chladni, who first observed and described them. Using a simple lapel microphone, these mode patterns can also be illustrated by noting the phase shift when observing the near-field sound pattern very close to the vibrating plate. Both Chladni patterns and phase mapping of patterns will be illustrated. A relatively inexpensive software package enables finite element calculations of vibrating plates. Normal mode vibration can easily be illustrated in animation on a computer screen. Both the calculation process and the animation will be illustrated.

9:30

3aED5. Acoustics outreach and scouting: A merit badge proposal. Kent L. Gee, Brian E. Anderson, Tracianne B. Neilsen, and Scott D. Sommerfeldt (Dept. of Phys. and Astron., Brigham Young Univ., N283 Eyring Sci. Ctr., Provo, UT 84602, kentgee@byu.edu)

The success of the Girl Scouts of America (GSA) and the Optical Society of America's "Project Lighten Up" and the proposed "Project Listen Up" have prompted a discussion of how Scouting can be used as an outreach vehicle to interest today's youth in science and, more particularly, in acoustics. One possibility for this would be the drafting of a "Sound and Audio" merit badge for the Boy Scouts of America (BSA) or a similar activity award for the GSA. The BSA currently offers 121 merit badges for scouts, including varied topics such as archeology, rifle shooting, plumbing, truck transportation, and coin collecting. Although there are physical science-related merit badges (e.g., composite materials, chemistry, nuclear science), there is not a badge that deals directly with acoustics. This talk will outline how an acoustics-related merit badge may help engage youth interest in science and discuss ideas for possible badge requirements.

9:45

3aED6. Wave interaction effects for all ages. Scott D. Sommerfeldt (Dept. of Phys., Brigham Young Univ., N181 ESC, Provo, UT 84602, scott_sommerfeldt@byu.edu)

Wave interaction effects occur in many everyday situations. A number of important principles associated with these interactions can be readily demonstrated without the need for extensive and expensive equipment. Using two low-cost speakers, it is possible to easily demonstrate the basic principle underlying active noise control by placing the speakers close together and first wiring the speakers "in-phase" to achieve constructive wave interference and then wiring the speakers "out-of-phase" to achieve destructive wave interference. This setup results in a global attenuation of the sound field, making the audible effect easily discernible. The dependence of the interaction on frequency and spacing (kd) can be demonstrated by changing the frequency of the excitation signal. In addition, by separating the speakers the interaction of two waves to create a standing wave can be demonstrated, and the nodes/antinodes established can be audibly heard as the listener moves his/her ear through the field. Finally, these concepts can be used to understand why loudspeakers are mounted in enclosures. As the loudspeaker is placed in an enclosure the direct interaction of (out-of-phase) waves from the front and back of the loudspeaker cone is inhibited, thus resulting in increased radiation efficiency.

Contributed Papers

10:00

3aED7. A simple analytical demonstration of transient and steady-state response in a dissipative closed waveguide based on the d'Alembert solution. Jerry H. Ginsberg (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405, jerry.ginsberg@me.gatech.edu)

Previously [J. H. Ginsberg, J. Acoust. Soc. Am. **119**, 1954–1960 (2006)], the d'Alembert solution of the wave equation was adapted to address the response to initial conditions of a one-dimensional waveguide whose terminations have an arbitrary real (possibly dissipative) impedance. The technique creates virtual images on the other (i.e., nonphysical) side of each wall to convert the closed waveguide to one of infinite extent. The present work extends the development to cases where an arbitrary velocity is applied to one wall. The imaging technique allows one to visually track the evolution of the spatial pressure profile and the relation of temporal waveforms at various locations. The relatively low level of mathematical skills entailed in the formulation makes the method accessible to students at all levels. Application of the technique to periodic excitation of the boundary leads to visual representation of various effects, including reverberation time and various resonance effects, such as uncontrolled growth in the absence of dissipation and approach to steady-state when dissipation is present. Sound files generated from the synthesized connect the physical and audible phenomena.

10:15—10:30 Break

10:30

3aED8. Hands on and homemade coupled oscillators for Project Listen Up. Sean M. Mock and Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402, m114518@usna.edu)

A coupled pendulum oscillator experiment recently developed within the U.S. Naval Academy Physics Department seems a possible candidate for the Project Listen Up initiative. A plexiglass frame supported by four legs suspends two 500 g masses 30 cm below by two chains of paper clips each, meeting in the shape of a V on either bob. A light spring (k=3.64 N/m) connects the two, completing the coupled oscillator system. Displacement of the oscillators was measured using Pasco motion sensors, a Vernier LabPro unit, and a laptop. LOGGERPRO and MATHEMATICA software helped compare theory with experiment. However, one does not need these resources to experiment with coupled oscillators—it can be improvised using ordinary household materials for young students! An inexpensive wooden picture frame with islet screws supports pendulums using two Dixie cups filled with sand or coins as masses. The possibilities for learning seem endless! A simple stopwatch, meter stick, fulcrum, and measuring cups are incorporated in experimental and theoretical comparisons, including the periods of normal modes. Students can explore density through balancing masses with volumes of water. Demonstrating beat frequencies from mistuned xylophone bars or guitar strings conveys a subtlety of the experiment without overpowering the young minds of future scientists.

10:45

3aED9. Helmholtz resonator experiment for Project Listen Up. Marianne C. Sparklin and Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402, m096120@usna.edu)

The behavior of sound waves and resonant effects can be observed using Helmholtz resonators. Resonators are built from identical round wooden boxes purchased from a craft store. The tops are glued and drilled with hole diameters from 2–3 cm. We can study the resonant behavior as a function of hole diameter only. The end effect is large compared with the thickness of the orifice (0.6 cm). In the experiment, a swept tone (100–900 Hz) from a small speaker drove the resonator. One studied the output of the microphone located near the orifice. One could use a swept spectrum analyzer, a function generator, or tuning fork with varying weights to measure the resonant frequency of the box. A plot of hole diameter versus frequency can be compared with the theoretical Helmholtz resonant frequency prediction, which depends on the volume of the box, the cross sectional area, the sound speed of air, and the effective length of the hole. One can model the effective length of the resonator as the thickness of the top plus a constant times the diameter of the hole and get good agreement with the established end correction found in textbooks.

11:00

3aED10. The energy flow for a spherical acoustic lens: Ray and wave methods vs experiment. Cleon E. Dean (Phys. Dept., P.O.B. 8031, Georgia Southern Univ., Statesboro, GA 30460-8031, cdean@GeorgiaSouthern.edu) and James P. Braselton (Georgia Southern Univ., Statesboro, GA 30460-8093)

A simple classroom demonstration consists of a weather balloon filled with carbon dioxide, a sound source, and a microphone. Since the speed of sound is slower in carbon dioxide than in air at room temperature and pressure, the balloon acts as a positive spherical acoustic lens. The accuracy of ray methods in locating the acoustic focus versus a full-blown wave solution approach has been presented previously [Cleon E. Dean and James P. Braselton, "The energy flow for a spherical acoustic lens: ray vs. wave methods," J. Acoust. Soc. Am. **123**, 3520 (2008).] Theoretically, this problem presents particular difficulties if the sound source lies in the near field region. The sound emitter is treated as a dipole source equivalent to a rigid oscillating sphere of small size and amplitude of motion relative to the scatterer. The energy flux around the balloon has been visualized by both ray methods and by the acoustic energy flux vector field. The geometrical ray results and the energy flux vector field resulting from the wave solution are compared with experiment.

11:15

3aED11. The effect of the teaching-learning sequences connecting sound wave at microlevel and macrolevel on the student's understandings for sound transmission. Hyungjun Kim (Phys. Education Dept., Seoul Natl. Univ, Daehak-dong, Gwanak-gu, Seoul 151-742, che94@dreamwiz.com) and Junehee Yoo

Previous studies reported that students had some difficulties in understanding sound waves as longitudinal waves. In this research, most of teaching-learning sequences described sound waves at macrolevel, which were supposed as one source of these difficulties. So teaching-learning sequences connecting sound wave at microlevel and macrolevel were developed and effects were investigated. These teaching-learning sequences were focused on bridging between the movement of particles in air and movements of slinky springs and simulation using the EXCEL program. In addition to that, explaining wave phenomena at macrolevel by the movement of particles at microlevel was emphasized. Predict-observe-explain strategy was used. Evaluated worksheets and video-records were analyzed to figure out conceptual changes.

11:30

3aED12. Modal analysis for undergraduate laboratories and projects. Andrzej Kotlicki and Chris Waltham (Dept. of Phys. and Astronomy, Univ. of British Columbia, 6224 Agricultural Rd., Vancouver, BC V6T 1Z1, Canada)

Vibrational analysis of structures (musical instruments for example) requires a means of excitation, motion detection and a data acquisition system. None of these needs to be very expensive. We have constructed an impact hammer using a piezoelectric crystal from an old barbeque lighter; this provides rapid excitation at all frequencies up to 1 or 2 kHz. Motion detection is now possible using extremely light (less than 1 g) microelectromechnical systems (MEMS) accelerometers that cost only a few dollars each. Two-channel data acquisition at 44.1kHz per channel is available to anyone with a computer equipped with a soundcard. More flexible external systems are also available in the $200 range. Thus, modal analysis is easily within the budgets of undergraduate laboratories. In this presentation we will emphasize the calibration of the impact hammer and evaluation of the MEMS accelerometers.

Session 3aNS

Noise: Noise Litigation

John Erdreich, Chair
Ostergaard Acoustical Assoc., 200 Executive Dr., West Orange, NJ 07052

Chair's Introduction—8:00

Invited Papers

8:05

3aNS1. Hearing loss litigation. Robert A. Dobie (Dept. of Otolaryngol., UC-Davis, Sacramento CA 95817)

Most hearing loss in American adults is attributable to aging, noise, or a combination of these two. While most claims for noise-induced hearing loss are handled administratively in workers compensation, litigation against employers, premises owners, and manufacturers of noisy equipment continues to pose interesting challenges. Issues of negligence often involve professionals with expertise in noise control, hearing conservation program management, product labeling, and human factors. There are usually two primary issues for clinician experts: How severe is the hearing loss, and what caused it? Audiologists and otologists can draw upon a mass of research to support reasonable opinions in cases with complete data sets. Unfortunately, important data from the past (e.g., changes in noise exposure and hearing loss over a career) are often impossible to obtain. Forensic consulting is terra incognita for most clinicians, requiring analyses and judgments that are rarely called for in clinical practice.

8:35

3aNS2. Minneapolis et al versus the Metropolitan Airports Commission. Paul Schomer (Schomer and Assoc., Inc., 2117 Robert Dr., Champaign, IL 61821, schomer@SchomerAndAssoc..com)

The city of Minneapolis has a law protecting the quietude in various land use areas, including residential areas. In this legal action, plaintiff asserted that the Metropolitan Airports Commission was adversely and significantly affecting the quietude of the residents in the 60–65 DNL zone and that a 5-dB noise insulation package in homes in that zone would ameliorate the situation. There were four tasks for plaintiff's expert witnesses; (1) Define quietude; (2) Show that without the airport quietude existed in the 60–65 DNL zone; (3) Quantify the impacts including annoyance, speech interference, and sleep disturbance; and (4) Quantify the benefits of a 5-dB noise insulation package. The results were a negotiated settlement for $128 million.

9:05

3aNS3. Qualification of an aircraft ground run-up enclosure using a novel sound monitoring and analysis approach. Kerrie G. Standlee (Daly-Standlee & Assoc., Inc., 4900 SW Griffith Dr., Ste. 216, Beaverton, OR 97005)

This paper describes a sound monitoring and analysis program developed to determine if a ground run-up enclosure (GRE) installed at the Portland International Airport was effective in reducing aircraft engine run-up noise radiated to residential neighborhoods during nighttime hours. A noise monitoring and analysis program was developed to fulfill conditions of approval imposed by the Oregon Department of Environmental Quality (DEQ) prior to the construction of the facility. A novel approach was developed whereby unattended sound level meters were used to collect one-second Leq and hourly statistical noise level data near the GRE and at community locations of interest. This data, combined with GRE usage data and weather data, was then subjected to a multistep analysis procedure. The process was found to be very effective at determining when ground run-up events caused community noise levels exceeding DEQ limits. By the end of the project, the procedure was being used to assess 99 percent of the run-ups occurring within the GRE. The sound monitoring and analysis procedure became a very reliable tool in demonstrating that the GRE complied with its conditions of approval.

9:35

3aNS4. Noise issues affecting litigation in condominiums. Jerry G. Lilly (JGL Acoust., Inc., 5266 NW Village Park Dr. Issaquah, WA 98027)

Serving as an expert witness in legal actions involving residential condominiums is often a harrowing and extremely challenging task, regardless of which side you represent. If you are representing the HOA, you have to educate your client so that they understand that the mere ability to hear your neighbor is not necessarily unusual and may not be grounds for a defect. If you are representing the architect, developer, or contractor, you have to educate them as to the code requirements as well as the various acoustical performance guidelines that will undoubtedly be introduced by the HOA. In either event, if you do not do your homework you are leaving yourself exposed to be grilled by opposing counsel (in deposition or at trial) without mercy. This presentation will discuss the author's experience in dealing with these issues, with specific emphasis on problems with the testing standards and performance guidelines that exacerbate the problem.

10:05—10:20 Break

10:20

3aNS5. Warning audibility in industrial accidents. John Erdreich (Ostergaard Acoust. Assoc., 200 Executive Dr., W. Orange, NJ 07052)

Industrial accidents related to warning audibility may be adjudicated in the worker compensation environment or in state courts. One factor that may determine whether the incident becomes a compensation case or a tort is if the employer could reasonably expect the situation to require an audible warning. The second factor to evaluate is the adequacy of the warning itself. Thus, it is the responsibility of the acoustician to educate the trier of fact about not only the science of audition, but also the state of the art in that represents a consensus of methods to provide such warnings. An example of such a case will be presented.

Contributed Papers

10:50

3aNS6. Can race car sound break windows? David Braslau (David Braslau Assoc., Inc., 1313 5th St. SE, Ste. 322, Minneapolis, MN 55414)

Sound from a special class of race cars was cited as cause for sudden breakage of a large pane of tempered glass in a private box that fell and injured a spectator below. Analysis of previously recorded race track noise from a similar track in Minnesota showed that sound levels could exceed the Hubbard rattle criteria but no conclusions could be drawn regarding breakage. Because of time constraints, sound level measurements from a race in Ohio of this vehicle class were taken by Campanella Associates for reference. Those measurements were compared with previously measured levels of other vehicle classes and found to be higher at some frequencies. Since a recorded demonstration performed by the plaintiff showed breakage of a similar pane of glass with an induced flaw and continuously applied sound at 70 to 80 Hz, a field test on a similar window with simulated sound levels based upon measurements from the special vehicle class race was performed with assistance from Synergistic Design Associates. A simulated sound level as high as 117 dBA was generated at the window without breakage. Some conclusions on potential glass breakage from incident sound will be presented. [Work supported by Johnson & Lindberg, P.A.]

11:05

3aNS7. Deficiency of the Massachusetts Pure Tone Noise Regulation. Michael Bahtiarian (Noise Control Eng. Inc., 799 Middlesex Trnpk., Billerica, MA 01821, mikeb@noise-control.com)

The Commonwealth of Massachusetts Department of Environmental Protection (MADEP) noise regulation has two parts. The first part involves overall A-weighted noise and requires that no source of sound be greater than 10 dB above the background noise. A second part states that no pure tones are allowed. A pure tone condition is defined by the MADEP as any octave band that is greater than the adjacent bands by 3 dB. The author was involved in community noise investigation that showed the deficiency of the pure tone requirement. A suburban community complained of bothersome noise from an industrial abutter. Noise measurements were taken by other consultants which did not register as a pure tone. The author was hired by the community group to conduct measurements on their behalf. Similar octave band measurements were taken with similar results. However, measurements with finer one-third octave band showed significant tonal activity. When compared to commercial standards for pure tones the tonal criterion was met. Further examination of narrowband data showed why the original octave band measurements were not registering pure tone conditions. The author will present his suggestion for changes to the MADEP regulation and welcomes input from others dealing with similar tonal requirements.

11:20

3aNS8. Uncertainty in field and laboratory acoustical testing: A challenge for the acoustical industry. John LoVerde and Wayland Dong (Veneklasen Assoc., 1711 16th St, Santa Monica, CA 90404, jloverde@veneklasen.com)

The statistical uncertainty in acoustical testing has been insufficiently studied in the acoustical community. There are a limited number of studies quantifying measurement uncertainties, especially in the field, but even where uncertainties have been documented, they are largely ignored in practice. As was documented in a previous case [LoVerde and Dong, J. Acoust. Soc. Am. **124**, 2441 (2008)], it is common practice to assume that the test results have low uncertainties (even if published uncertainties are large) and that variations in the results are due to variations in materials, errors in construction, etc. This can be particularly important in noise litigation, where acoustical ratings that are lower than expected are often automatically assumed to be due to construction defect, even though they may be within the expected range. Even without errors in construction, the uncertainties in the acoustical test procedure may be such that some assemblies will meet the relevant statutory requirements and other, nominally identical assemblies will fail. Recent experience and insight will be presented regarding the definition of a construction defect in acoustical assemblies, the overlap of responsibilities between contractors and designers, and the interpretation of building code requirements.

11:35

3aNS9. Designing petrochemical plants to permit retrofitting of any additional controls needed to meet community noise limits. Frank H. Brittain (Bechtel Corp., 2255 Peavine Valley Rd., Reno, NV 89523)

Designing petrochemical, power, or industrial plants to meet community noise limits can be expensive. To meet stringent limits, incremental costs often exceed one million dollars per decibel. There is considerable uncertainty, including: overly conservative or unrepresentative vendor noise data, limitations of ray-tracing software, questionable accuracy beyond 1000 m, lack of TL and absorption data below 125 Hz, unreliable IL data for pipe lagging at all frequencies, unexpected operating conditions, and vendors not meeting their equipment noise limits. Atmospheric effects can cause variations of A-weighted community noise levels of 25 dB at 100 m from a refinery, which can make meeting a not-to-exceed noise limit difficult and risky. Further, lack of experience with noise control of any participant (noise control engineer, project, owner, regulators, or vendors) can substantially increase uncertainty. Including sufficient noise controls to cover these uncertainties will be expensive, often overly conservative, and often unnecessarily expensive. One way to substantially lower costs and risks is designing the plant to permit delaying some but not all controls until start-up—if needed. This paper describes the planning needed to permit a partial reliance on retrofit controls, what type of controls can be delayed until start-up, and space allocation. Examples are included.

Session 3aPA

Physical Acoustics: Infrasound and Outdoor Sound Propagation

D. Keith Wilson, Chair
U.S. Army Cold Regions Research Lab., Engineering Research and Development Ctr., Hanover, NH 03755-1290

Contributed Papers

8:30

3aPA1. On tropospheric ducting in infrasound. Roger Waxler and Philip Blom (NCPA, Univ. of Mississippi, University, MS 38677, rwax@olemiss.edu)

Infrasound can propagate to ranges of many hundreds of km. Signals received on the ground at ranges of 200 km or more have propagated high into the atmosphere through the stratosphere and thermosphere. At shorter ranges, the propagation is dominated by structure in the lower atmosphere such as the jet stream and the atmospheric boundary layer. Theoretical results on the influence of the lower atmosphere on infrasound propagation will be presented.

8:45

3aPA2. On the fast arrivals seen in impulse propagation in the nocturnal boundary layer. Philip Blom and Roger Waxler (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

On clear dry nights over flat land, an acoustic duct forms in the atmospheric boundary layer. The duct is formed by a temperature inversion in the lowest few hundred m of the atmosphere and by the stiff wind, known as the nocturnal jet, which commonly develops above the temperature inversion. An impulsive signal undergoes strong dispersion as it propagates in the nocturnal duct and is received at long ranges from the source as an extended wave train consisting of a series of distinct arrivals followed finally by a low frequency tail. The first arrival is of interest here. Through its interaction with the nocturnal jet, the first arrival is often anomalously fast. Accompanying the increased propagation speed is a marked decrease in amplitude. It will be shown that an explanation for this behavior is given by geometrical acoustics.

9:00

3aPA3. Impulsive sound propagation in coastal regions. Xiao Di, Carrick Talmadge, Roger Waxler, Kenneth E. Gilbert (Natl. Ctr. for Phys. Acoust., Univ. of Mississippi, University, MS 38677), and David A. Conner (U.S. Army Aviation and Missile Command)

Influenced by the coastal temperature profile and wind profile, the resulting sound speed profiles in coastal regions have certain unique features, such as double inversion zones. As a result, long range sound propagation presents certain properties due to these features. An impulsive sound propagation experiment had been conducted in the morning hours of the summer of 2007 in the coastal Florida area. The test field was 7 km long. Three propane cannons were used during the experiment, located at each end of the test field and in the middle of the test field. Both upwind and downwind propagation data were collected and will be presented. The data showed that coastal region meteorology has a dominant effect on long range sound propagation. Simulated sound propagation results using measured meteorological data are compared with measured impulsive sound propagation data.

9:15

3aPA4. Investigation of dominant sources of pressure fluctuations and infrasonic wind noise in microphones. John Paul Abbott, Richard Raspet, Jeremy Webster, and Jiao Yu (Natl. Ctr. for Physical Acoust., Dept. of Phys. and Astronomy, Univ. of Mississippi, University, MS 38677)

Investigation of the time dependent relationship between the vertical wind velocity fluctuations and the pressure sensed on the ground as measured in an open field is done to determine if the turbulence-shear term of the Poisson equation is the dominant source of pressure fluctuations and infrasonic wind noise present in microphones, or some other source. The wind velocity was measured using a three-axis ultrasonic anemometer and six one-dimensional ultrasonic anemometers. The pressure was measured simultaneously using two half-inch free-field pressure microphones, one placed flush with the ground, the other at a height of 1 m above the first. The anemometers were also situated above and adjacent to the microphones. The three axis anemometer was set at a height of 1 m, and the six, one-dimensional anemometers, as a vertical array from a height of 0.1–2 m above the ground. As this is still an ongoing investigation, only a description of the experiment and the current findings are presented and discussed.

9:30

3aPA5. New models for wind noise measured in a flat surface under turbulent flow. Jiao Yu, Richard Raspet, Jeremy Webster, and JohnPaul Abbott (Natl. Ctr. for Physical Acoust. and Dept. of Phys. and Astron., Univ. of Mississippi, University, MS 38677, jyu2@olemiss.edu)

We have previously developed models for predicting the power spectral density of the wind noise pressure measured in a flat plate outdoors from the measured power spectral density of the turbulence and the measured wind velocity profile above the plate [Yu *et al.*, Proceedings of NCAD 2008, NoiseCon2008-ASME NCAD]. Recently we have corrected an error in the model for the logarithmic profile wind velocity gradient results and have developed an improved integration method. Also, we have developed a prediction for arbitrary wind velocity profiles using the previous single exponential model. Typical results comparing our predictions with our measurements are presented and analyzed. A simple algebraic fit to the prediction for the logarithmic profile fit form is also provided for use by others. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

9:45

3aPA6. Calculations of wind noise underneath a foam covering at the ground surface. Jiao Yu, Richard Raspet, Jeremy Webster, and JohnPaul Abbott (Natl. Ctr. for Physical Acoust. and Dept. of Phys. and Astron., Univ. of Mississippi, University, MS 38677, jyu2@olemiss.edu)

A foam windscreen reduces the wind noise generated at a microphone. Similarly, a thin foam covering above a surface mounted microphone also lowers the wind noise level effectively. Recent research has focused on developing a physical model for calculating the wind noise pressure fluctuation spectrum measured by a foam covered microphone under a turbulent flow based on our current model for flush mounted bare microphones. The foam surface not only reduces the wind noise at the microphone, but also prevents direct exposure of the microphone to the flow, which reduces the measurement's sensitivity to details of the meteorological conditions. In this talk, the calculation model will be introduced, and the comparisons between our predictions with the model and our measurements down to low wave number will be presented and discussed. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

10:00—10:15 Break

10:15

3aPA7. Study of the impact of input uncertainties on acoustic propagation. Bruce Ikelheimer, Micah Downing, and Michael James (Blue Ridge Res. and Consultants, 13 1/2 W. Walnut St., Asheville, NC 28801)

A modular system of acoustical tools has been developed that integrates the physical factors controlling the propagation of acoustical signals. First and foremost, this system provides a prediction of the noise from a source to a receiver location. This ray-tracing prediction takes into account the local terrain, changes in ground impedance, turbulence, and thermal and wind gradients. This propagation prediction is combined with a noise source that can handle fully three-dimensional, spectral source descriptions. The model uses a six-dimensional definition of the source location and orientation to select the radiation angle from the source to the receiver. Levels of audibility of the received signal are then calculated based on the received noise levels, the local ambient noise levels, and the expected listener acuity. Each input to the model has uncertainties with the potential to generate errors that cascade through the system. Differing atmospheric, terrain, and ground impedance conditions have been tested in our model both individually and in combinations to see how they interact. The goal is to have an understanding of the confidence level of audibility provided a known level of uncertainty on the input data.

10:30

3aPA8. Acoustic temperature measurement in a full-scale rocket noise field. Jarom H. Giraud, Kent L. Gee, John E. Ellsworth, and Derek C. Thomas (Dept. of Phys. and Astronomy, Brigham Young Univ., N283 ESC Provo, UT 84602)

Acoustic pressure measurements in high-amplitude sound fields are common, but acoustic temperature measurements are relatively rare. This is primarily because thermometers exhibit thermal lag and, consequently, a limited frequency range. In this study, a 1 μ diam platinum wire resistance thermometer was used to measure temperature fluctuations generated by propagating noise produced by a horizontally fired, static GEM-60 solid rocket motor. The data are compared with those calculated from the pressure data of a nearby 3.18 mm condenser microphone by assuming an ideal adiabatic process.

10:45

3aPA9. Rise times of low-energy, free-space propagating acoustic shocks. Thomas B. Gabrielson, Timothy M. Marston (Graduate Program in Acoust., Penn State Univ., PO Box 30, State College, PA 16804), and Mark S. Wochner (The Univ. of Texas at Austin, Austin, TX 78713-8029)

Characteristics of the "ideal" shock, a one-dimensional step change in pressure, are reasonably well established; however, the freely propagating shock with finite duration is challenging to model and measure. These challenges are reflected in an extensive body of literature over several decades. In this paper we present a method for producing free-space propagating shocks, measurements of rise time with minimal contamination by turbulence, and model results that account for spreading, absorption (including molecular relaxation), and nonlinearity. The shocks are produced using an acoustic pipe filter to condition the irregular pressure release of either balloon pops or starter-pistol shots. The shocks range in peak pressure from 10 to 1000 Pa and in rise time from 10 to 0.5 μs. Rise-time measurements are made using a custom piezoelectric polymer wideband microphone with a measurement range that overlaps with that of a diffraction-corrected 1/8-inch measurement microphone. The exponential-decay time constant beyond the pressure peak is typically less than 100 μs, which may be too rapid to permit development of characteristics associated with relaxation mechanisms: rise times over the entire range measured here are well predicted by accounting for classical and rotational absorption only.

11:00

3aPA10. Three-dimensional numerical simulation of sonic boom diffraction around buildings using a dispersion-relation-preserving scheme. Sang I. Cho and Victor W. Sparrow (Grad. Program in Acoust., The Penn State Univ., 201 Appl. Sci. Bldg., Univ. Park, PA 16802)

A finite difference time domain approach is used to model the temporally and spatially dependent pressure loading on a rigid structure due to an incident low-amplitude N-wave sonic boom. Analytically, one expects strong diffraction due to building corners and pressure doubling near rigid walls. The three-dimensional linear wave equation is sufficient to model the interaction between the structure and the low-amplitude sonic boom at the ground level. Applying a fourth-order dispersion-relation-preserving scheme and a four-level explicit time marching scheme, an excellent dispersion characteristic is achieved without sacrificing computational efficiency of the algorithm. NASA flight testing at Edwards, CA during the summer of 2006 is numerically simulated by matching the geometries of the simulated structure to the actual residential house used for the flight test and using the waveform of a recorded boom as the incoming wave for the simulation. A direct comparison to the experimental data allows the quantitative validation of the simulation results. The feasibility of using this model for calculation of the pressure loads on other structures can be determined based on the accuracy of this numerical solution. [Work supported by NASA. The authors appreciate NASA making the 2006 and 2007 flight data available for this work.]

Session 3aPP

Psychological and Physiological Acoustics: Impacts of Hearing Impairment

Peggy B. Nelson, Chair
Dept. of Speech, Language and Hearing Science, Univ. of Minnesota, Minneapolis, MN 55455

Contributed Papers

8:45

3aPP1. A personal perspective on unilateral hearing loss. Brian D. Dushaw (Appl. Phys. Lab., Univ. of Washington, 1013 N.E. 40th St., Seattle, WA 98105, dushaw@apl.washington.edu)

Although I have been unilaterally deaf since birth, only later in life have I realized the extent of this handicap and how I can best manage it. The effects of this hearing loss are complicated, being an interplay of physics, psychology, social interaction, etc. Because people are able to adapt, the extent of this "unseen" handicap is underappreciated. Responses by the public to an informative website (http://909ers.apl.washington.edu/~dushaw/SSD) suggest that this handicap has a universal nature. Obvious effects are the inabilities to locate sound or to bilaterally process sound signals. Less obvious effects are a heightened sensitivity to noise: as masking to a desired signal, as a fatigue factor, as an impediment to mental concentration. The difficulties encountered when talking over the telephone to someone who is speaking in a noisy environment may give one an appreciation of the effects. Even less obvious effects are the long-term social implications of the handicap: the tendency to ignore those on one's deaf side, the inability to function "normally" at noisy events such as parties, ASA receptions, etc., or a tendency to irritability in noisy environments. In this unscientific talk, I will attempt to highlight issues that may be deserving of additional study.

9:00

3aPP2. Masking period patterns of exponentially ramped and damped noises in hearing-impaired listeners. Jennifer Lentz and Yi Shen (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN)

Noises that are modulated with either an exponentially rising (ramped), repeating modulator or a falling (damped), repeating modulator may evoke strong perceptual differences. To explore a possible cochlear contribution to these perceptual differences, masking period patterns (MPPs) of ramped and damped noises were measured for normal-hearing listeners at different stimulus levels and for hearing-impaired listeners. By time-reversing the ramped MPP, comparisons can be made between thresholds for which the masker power was the same between ramped and damped noises. Previous results showed that for normal-hearing listeners, the MPP of a ramped noise is considerably shallower than the MPP of a damped noise. The present results indicated that the differences between the damped and the time-reversed ramped MPPs were the largest at moderate stimulus levels and the smallest at low and high stimulus levels. Listeners with hearing loss exhibited no difference in the damped versus time-reversed ramped MPPs. Results will be discussed in terms of contributions of the cochlear nonlinearity and central auditory processes.

9:15

3aPP3. Modulation detection interference in listeners with cochlear hearing loss: Effect of modulation depth and onset delay. Jungmee Lee and Soonha Yook (Dept. of Speech, Lang. and Hearing Sci., Univ. of Arizona, 1131 E. 2nd St., Tucson, AZ 85721)

Modulation detection interference (MDI) was measured for listeners with normal hearing (NH) and cochlear hearing loss (HI). The signal was 1 kHz, and the interferer was 2 kHz. The stimuli level was 50 or 22 dB SL. The interferer was modulated with 8 Hz at various modulation depths (25, 50, 75, or 100%). The onset of signal was delayed by 0, 125, 250, 375, 500 or 625 ms relative to onset of the interferer. For NH subjects, MDI increased systematically with increasing modulation depth of the interferer. MDI almost disappeared with 375 or 500 ms onset delay except a condition with the interferer of 25% modulation depth where there was no MDI at all across different onset delay conditions. In contrast, for HI subjects, there was quite an amount of MDI even with the interferer of 25% modulation depth, and MDI was greater than NH for all modulation depths at 0 and 125 ms onset delays, suggesting that the perceived modulation depth of the interferer might be exaggerated for HI than for NH.

9:30

3aPP4. Frequency modulation detection, frequency selectivity, and speech recognition in noise by hearing-impaired listeners. Van Summers, Matthew Makashay, Joshua Bernstein (Army Audiol. & Speech Ctr., Walter Reed Army Medical Ctr., 6900 Georgia Ave., NW, Washington, DC 20307), Sarah Melamed, and Marjorie Leek (VA Medical Ctr., Portland, OR 97201)

Reduced ability to make use of temporal fine structure (TFS) information may contribute to deficits among hearing impaired (HI) listeners for speech recognition in competing background sounds. The broad tuning typically observed in HI listeners will produce more complex patterns of TFS as input to the auditory nerve, and these patterns may be more difficult to encode and/or interpret by higher auditory centers. In the current study, we tested HI listeners on a frequency-modulation (FM) detection task thought to require the use of TFS cues, on tone detection in notched-noise to assess frequency selectivity, and on speech recognition in continuous and modulated background sounds. Performance on the psychoacoustic tasks was tested at four frequencies (500, 1000, 2000, and 4000 Hz) at levels between 60 and 85 dB SPL (approximating the range of third-octave band levels in the speech stimuli). HI listeners with similar audiometric thresholds showed clear differences in frequency tuning, FM detection thresholds, and speech performance. Relationships between the psychoacoustic measures in the four frequency regions and between these measures and speech performance will be discussed.

9:45

3aPP5. Relative contributions of spectral and temporal cues for speech recognition in patients with sensorineural hearing loss. Li Xu, Ning Zhou, Katherine Rife, and Rebecca Brashears (School of Hearing, Speech and Lang. Sci., Ohio Univ., Athens, OH 45701)

The present study was designed to examine speech recognition in patients with sensorineural hearing loss (SNHL) when both temporal and spectral information in the speech signals were co-varied. Four subjects with SNHL were recruited to participate in consonant and vowel recognition tests that used speech stimuli processed through a noise-excited vocoder. The number of channels was varied between 2 and 32, which defined spectral information. The lowpass cutoff frequency of the temporal envelope extractor was varied from 1 to 512 Hz, which defined temporal information. Results indicate tremendously individual differences among the subjects with SNHL. For consonant recognition, patterns of relative contributions of spectral and temporal information were similar to those in normal-hearing subjects. The utility of temporal envelope information appeared to be normal in the hearing-impaired listeners. For vowel recognition, the performance plateau was achieved with numbers of channels as high as 16–24, much higher than expected, given that the frequency selectivity in patients with SNHL might be compromised. In order to understand how hearing-

impaired listeners utilize spectral and temporal cues for speech recognition, future studies will be necessary to elucidate the relationship between frequency selectivity as well as central processing capability and speech recognition performance using vocoded signals.

10:00—10:30 Break

10:30

3aPP6. Discrimination of formant-like glides in gated noise: Effect of sensation level. Peggy Nelson, Yingjiu Nie (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, peggynelson@umn.edu), Magdalena Wojtczak, and Gordon Stecklein (Univ. of Minnesota, Minneapolis, MN 55455)

Listeners with sensorineural hearing loss (SNHL) do not experience masking release for speech in fluctuating noise, as normal-hearing listeners do. One hypothesis suggests dip listening may be difficult because frequency glides, or speech formant transitions, may be difficult to discriminate at low sensation levels when in the presence of fluctuating noise. Normal-hearing listeners were tested for their ability to discriminate 100-ms frequency glides with center frequencies of 800 and 2500 Hz. They were tested in quiet and in the presence of 500-ms of 8-Hz square-wave gated broadband noise at a signal-to-noise ratio of −10 dB. Listeners were tested at low (10–15 dB SL) and moderate (30–40 dB SL) signal levels. Preliminary results suggest that glide discrimination thresholds (measured as % change) are similar for low- and moderate-level stimuli in quiet; glide thresholds are poorer in gated noise than in quiet; and glide thresholds are poorest when low-level stimuli were presented in gated noise. These results suggest that when formants or glides are presented at low levels and in gated noise, discrimination abilities may be reduced. Implications for understanding masking release for speech by listeners with hearing loss will be discussed. [Work supported by NIH 5R01DC008306.]

10:45

3aPP7. Comodulation masking release and speech perception: Implications for dip-listening by cochlear implant patients. Antje Ihlefeld (MRC Cognition and Brain Sci. Unit, 15 Chaucer Rd, Cambridge CB2 7EF, England, antje1@gmail.com), Barbara G. Shinn-Cunningham (Boston Univ., Boston, MA 02215), and Robert P. Carlyon (MRC Cognition and Brain Sci. Unit, Cambridge CB2 7EF, England)

Compared to normal-hearing (NH) listeners, cochlear implant (CI) listeners struggle to identify speech when a source of noise interferes and do poorly when trying to listen in the dips of a modulated masker. In NH listeners, masker energy outside the spectral region of a target tone can improve target detection, a phenomenon referred to as comodulation masking release (CMR). Here we examined whether a comodulated flanking noise improves speech identification in a modulated noise masker, both for NH listeners identifying vocoded speech and for CI listeners. In experiment 1, NH listeners identified noise-vocoded speech (taken from a closed-set speech corpus) in a background of speech-shaped noise with or without a flanking narrow band of noise outside the spectral region of the target. The masker and flanker were either both 16-Hz square-wave modulated or were unmodulated. Performance was better in modulated than in unmodulated noise and, importantly, this difference was greatest when the comodulated flanker was present, consistent with the idea that CMR can introduce segregation cues between target and masker that help to listen in the dips of the masker. Experiment 2 tested CI listeners with the same experimental paradigm.

11:00

3aPP8. Masking release for speech in listeners with real and simulated hearing impairment. Joseph G. Desloge, Charlotte M. Reed, Louis D. Braida, Zachary D. Perez, and Lorraine A. Delhorne (Res. Lab. of Electron.s, Massachusetts Inst. of Technol., 77 Massachusetts Ave., Cambridge, MA 02139, jdesloge@mit.edu)

The current study examined the effects of audibility and age on the release of masking for speech in interrupted versus steady-state noise in listeners with real and simulated hearing loss. The absolute thresholds of each of ten hearing-impaired listeners were simulated in normal-hearing listeners through a combination of spectrally shaped masking noise and multiband expansion for the octave bands with center frequencies from 0.25–8 kHz. Each individual hearing loss was simulated in two groups of three normal-hearing listeners (one age-matched group and one non-age-matched group). The speech-to-noise ratio (S/B) for 50%-correct identification of HINT sentences was measured in backgrounds of steady-state noise and temporally modulated (10-Hz square-wave) noise at two overall levels for unprocessed speech and for speech that was amplified with the NAL-PR prescription. Results indicate that the magnitude of the release from masking (the difference in S/B obtained in steady-state versus interrupted noise) observed in individual hearing-impaired listeners was generally well produced in both groups of simulated-loss listeners. Thus, release of masking appears to be determined primarily by audibility regardless of age. Predictions of masking release derived from the Speech Intelligibility Index (SII; ANSI S3.5-1997) will be compared to the observed values. [Work supported by NIH-NIDCD R01 DC00117.]

11:15

3aPP9. Spectral modulation detection training in older adults with hearing loss. Andrew T. Sabin, Cynthia A. Clark (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr. Evanston, IL 60201, a-sabin@northwestern.edu), David A. Eddins (Univ. of Rochester, Rochester, NY 14618), Sumitrajit Dhar, and Beverly A. Wright (Northwestern Univ., Evanston, IL 60201)

Individuals with hearing loss are poorer than normal-hearing listeners at detecting patterns of peaks and valleys of sound level spread across audio frequency (spectral modulation detection)—an important skill for speech and music perception. In this preliminary report we examined whether listeners with hearing loss could improve on spectral modulation detection with practice. We trained older adults with sensorineural hearing loss ($n=7$) ~1 hr/day for 7 days to distinguish a 400–3200 Hz noiseband with a flat-spectrum from one with a 2 cyc/oct sinusoidal spectral modulation. Modulation depth was varied adaptively to determine the detection threshold. Their mean thresholds improved significantly from 18.9 to 12.8 dB. These listeners also improved at both a lower and a higher untrained spectral modulation frequency (1 and 4 cyc/oct), but showed no change on a measure of frequency selectivity. Interestingly, a separate group of younger normal-hearing adults who received the same training ($n=7$) did not improve on the trained (6.9 to 6.6 dB) or untrained conditions. Normal-hearing listeners had lower pre-training thresholds than individuals with hearing loss, and across all listeners the magnitude of individual improvement was correlated to pre-training threshold ($r=0.92$). Listener age and/or hearing loss could have contributed to these across-population differences.

11:30

3aPP10. Multidimensional perceptual scaling of musical timbre by hearing-impaired listeners. Kelly Fitz, Matt Burk, and Martin McKinney (Starkey Labs, 6700 Washington Ave., S., Eden Prairie, MN 55344, kelly_fitz@starkey.com)

We examine the impact of hearing loss and hearing aid processing on the perception of musical timbre. Our objective is to identify significant timbre cues for hearing-impaired listeners, and to assess the impact of hearing aid signal processing on timbre perception. Hearing aids perform dynamic, level-dependent spectrum shaping that may influence listeners' perception of musical instrument timbres and their ability to discriminate among them. Grey ["Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. **61**, 1270 (1977)] showed that sustaining instrument tones equalized for level, loudness, and duration are distinguished primarily along three perceptual dimensions that are strongly correlated with the acoustical dimensions of: (1) spectral energy distribution, (2) spectral fluctuation, and (3) precedent high-frequency, low-amplitude energy. Following the work of Grey, we ask listeners having mild to moderately severe sensorineural hearing loss to rate pairs of synthetic musical instrument tones according to dissimilarity in aided and unaided conditions. We analyze the dissimilarity judgments to identify acoustical correlates for the significant dimensions of timbre discrimination. This analysis helps us to objectively estimate the impact of hearing aid signal processing on the ability of hearing-impaired lis-

teners to discriminate among musical instruments, an important component of musical listening.

11:45

3aPP11. The perceptual tone-noise ratio of iterated rippled noise in listeners with hearing loss. Marjorie R. Leek, Sarah Melamed, Michelle R. Molis, and Frederick J. Gallun (Natl. Ctr. for Rehab. Aud. Res., Portland VA Medical Ctr., 3710 SW US Veterans Hospital Rd., Portland, OR 97239)

Previous work has demonstrated that iterated rippled noise (IRN) produces a weaker pitch in hearing-impaired (HI) listeners than in normal-hearing (NH) subjects (e.g., Leek and Summers, 2001). In this study, the hypothesis was tested that the weaker pitch is a result of an increased perceptual noisiness of IRN for HI subjects. The tone-noise ratio (T/N) was developed by Patterson *et al.* (1996) to describe the relationship between the strength of the tone component and the strength of the noise component for a given IRN. The T/N was measured using an adaptive procedure to find the tone-plus-noise level that matched each IRN stimulus. Four NH and 4 HI subjects took part in the study. The stimuli were IRNs with delays ranging from 4 to 16 ms and with one to eight iterations. There was a significant effect of delay and of iteration number, with longer delays and more iterations producing greater T/N. There were no significant differences between subject groups. These findings suggest that the weaker complex pitch observed in listeners with hearing loss is not due to an abnormally noisy IRN percept but must be attributed to other impaired processing mechanisms. [Work supported by NIH.]

WEDNESDAY MORNING, 20 MAY 2009 FORUM SUITE, 8:30 A.M. TO 12:00 NOON

Session 3aSA

Structural Acoustics and Vibration: General Structural Acoustics

Rudolph Martinez, Chair
CAA-Alion Corp., 84 Sherman St., Cambridge, MA 02140

Contributed Papers

8:30

3aSA1. Scan-based near-field acoustical holography on partially correlated fields: Theory and numerical experiments. Michael D. Gardner, Kent L. Gee, Alan T. Wall, Scott D. Sommerfeldt (Dept. of Phys. and Astronomy, Brigham Young Univ., N283 ESC, Provo, UT 84602), Daniel Manwill, and Jonathan D. Blotter (Brigham Young Univ., Provo, UT 84602)

Conventional near-field acoustical holography (NAH) requires a coherent field. For a coherent source, a scan-based approach can be used with one reference microphone to stitch the phase. For a noncoherent field, scan-based NAH can be performed if the virtual coherence technique is used. This technique uses multiple reference microphones to decompose the partial field into mutually uncorrelated partial fields, which are then processed by some NAH method. In particular, the statistically optimized near-field acoustical holography (SONAH) method is used with modified Tikhonov regularization. Numerical experiments are performed on a series of point sources with source strengths chosen to mimic the source characteristics of high-powered jets. The experiments are designed to aid in determining the proper number and location of reference microphones for doing NAH work on high-powered jet noise. [Work supported by Blue Ridge Research and Consulting and the Air Force Research Laboratory.]

8:45

3aSA2. Scan-based near-field acoustical holography on partially correlated fields: Laboratory experiments. Alan T. W, Michael D. Gardner, Kent L. Gee, Scott D. Sommerfeldt (Dept. of Phys. and Astron., Brigham Young Univ., N283 ESC, Provo, UT 84602, alantwall@gmail.com), Daniel Manwill, and Jonathan D. Blotter (Brigham Young Univ., Provo, UT 84602)

Forms of near-field acoustical holography (NAH) can be applied to source reconstructions of arbitrarily correlated extended sources, such as jet exhaust plumes. Prior to implementing NAH to reconstruct the jet noise source region of a military jet aircraft, laboratory experiments are being performed. These experiments utilize statistically optimized NAH (SONAH), coupled with the virtual coherence method, to reconstruct the radiation from correlated, partially correlated, and uncorrelated loudspeakers. Results from these experiments will be shown. Reconstructions of this type require the use of multiple reference microphones. However, little is known about how many should be used or where they should be placed. Successful reference placement strategies are being explored and commentary on these results will also be given. [Work supported by Blue Ridge Research and Consulting and the Air Force Research Laboratory.]

9:00

3aSA3. Near-field acoustic holography in conical coordinates. Tracianne B. Neilsen, Kent Gee, and Michael Gardner (Dept. of Phys., Brigham Young Univ., Provo, UT 84602, tbn@byu,edu)

Near-field acoustical holography (NAH) techniques can be optimized if the method capitalizes on the geometry of the noise source under investigation. Helmholtz-equation least squares method (HELS) uses the solutions of Helmholtz equation in spherical coordinates as basis functions for the pressure field. HELS is an efficient NAH technique if the source and the measurement surfaces are spherical in nature. For nonspherical cases, such as radiation from a plate or bar, it takes a large number of functions to represent the field. In these cases, there is also a question about where to place the origin of the wave functions. In search of a HELS-type method that could be applied to nonspherical sources, a study into the features of conical coordinates has been conducted. Because the Helmholtz equation is separable in conical coordinates, the solutions can be used, in a manner similar to HELS, as basis functions to represent the pressure field. For conical coordinates, the basis functions are spherical Hankel functions and Lame functions. This HELS-type formulation in conical coordinates could be a natural choice for NAH on conical sources and may be appropriate for NAH on jet noise. [Work supported by Blue Ridge Research and Consulting and Air Force Research Laboratory.]

9:15

3aSA4. Determination of ultrasonic vibration modes of a graded solid cylinder using a modified wave-expansion technique. Betsabe Manzanares-Martinez (Div. de Ciencias e Ingeniería, Unidad Regional Sur de la Universidad de Sonora, Boulevard Lzaro Crdenas 100, Navojoa, Sonora 85880, Mexico, betsabe@manzanares.uson.mx), Jin-Yeon Kim (Georgia Inst. of Technol., Atlanta, GA 30322-0355), and Arturo Baltazar (CINVESTAV-Saltillo, Ramos Arizpe, Coahuila 25900, Mexico)

Research on wave propagation in cylinders—hollow, solid, or graded—has been driven by its potential use in the inspection of long-span

engineering structures. Its practical application largely depends on the understanding of modal vibration inside the structure. Typically, the plane wave approach is used to model phase velocity, group velocity and dispersion curves in periodic systems. This proves cumbersome when used to model finite periodic structures, semifinite plates, or concentric cylinders. Here, a simpler approach to determine the vibration modes and dispersion curves of a solid concentric cylinder using a modified wave-expansion approach, used mostly in phononic crystals research, is proposed. The advantage of this technique is the possible application on a wide variety of cylindrical geometries and graded structures with various transverse sections. The results of the proposed technique are compared with those obtained by the resonance scattering theory for the simple case of a solid cylinder with a single core having different mechanical properties. [Work partially sponsored by SEP-CONACYT through project # 58951.

9:30

3aSA5. Three component time reversal imaging using nonlinear elasticity. Timothy J. Ulrich, II, Pierre-Yves Le Bas, Robert A. Guyer (Los Alamos Natl. Lab., EES-17, MS D443, Los Alamos, NM 87545, tju@lanl.gov), Brian E. Anderson (Brigham Young Univ., Provo, UT), Michele Griffa (EMPA, Duebendorf, Switzerland), and Koen Van Den Abeele (Katholieke Univ. Leuven Campus Kortijk, Kortrijk, Belgium)

The time reversed elastic nonlinearity diagnostic (TREND) has been shown to have the ability to image near surface nonlinear scatterers (e.g., cracks and disbonds) using the normal (out of plane) component of motion. This technique has been advanced with the addition of three-component vector wave field detection using laser vibrometry. This advancement provides the ability of TREND to detect and image features of various orientations, increasing its capability and decreasing the likelihood of overlooking features due to the inability to excite them. Furthermore, the ability to selectively focus each component of motion provides the means to investigate the mechanism of the nonlinear wave generation at the nonlinear scattering sites. Experimental data are presented from samples containing known defects (location and orientation) confirming the sensitivity of three-component time reversal imaging in both linear and nonlinear ultrasonics to feature orientation.

9:45

3aSA6. Parameter study for Structural Health Monitoring based on ambient noise cross-correlation. Najib Abou Leyla, Emmanuel Moulin, Jamal Assaad, Sebastien Grondel, and Christophe Delebarre (UVHC, IEMN-DOAE, Le Mont Houy, 59313 Valenciennes, France)

The possibility of performing Structural Health Monitoring in aeronautic structures using ambient noise cross-correlation techniques has been investigated in previous works. The applicability of such methods has been validated by checking the reproducibility and the potential in terms of damage detection. In this paper, a more detailed description of the acoustic phenomena involved is given. The influence of the different parameters (source characteristics, sensors and damage positions, etc.) on the cross-correlation function has been studied. A theoretic study is presented and used to develop a simple simulation tool based on multipath reflections of ballistic waves. The simulated results have been successfully compared to finite element results in simple cases. This tool is then exploited to characterize and quantify the influence of the significant parameters. Finally, some optimization solutions of the technique are proposed, and experimentally validated.

10:00—10:15 Break

10:15

3aSA7. Detection of structural faults in pipelines with time reversal. Nicholas A. O'Donoughue, José M. F. Moura (Dept. of Elect. & Comput. Eng., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15217, nodonoug,moura@ece.cmu.edu), and Yuanwei Jin (Univ. of Maryland Eastern Shore, Princess Anne, MD 21853)

The volatile nature of natural gas makes it extremely important to ensure that transport and distribution pipelines remain free from defects, as leakage can result in explosions. Current methods for testing buried pipelines rely on periodic excavation of a section of pipe and attachment of large acoustic or magnetorestrictive sensors. These systems, while reliable, suffer from a high cost-per-test ratio. To reduce the power requirements, we consider ultrasound embedded devices that could monitor the pipelines continuously. The problem with ultrasound is that, with buried pipes, the propagation is highly dispersive. We propose to use time reversal (TR), a signal processing technique, to refocus the signals and use, to our advantage, the dispersive nature of the environment. This paper will focus on the modes generated by various acoustic probing signals and the echoes received with and without TR. We argue that TR will be most beneficial when there are several dispersive modes present, a scenario avoided in conventional techniques. We will present simulation results for the detection of faults in a pipeline using PZFlex.

10:30

3aSA8. Acoustic detection of bolt detorquing in structures. Joe Guarino (Dept. of Mech. and Biomedical Eng., Boise State Univ., Boise, ID 83725-2075, jguarino@boisestate.edu) and Robert Hamilton (Boise State Univ., Boise, ID 83725-2075)

Recent interest in structural health of bridges and large buildings has lent new urgency to the development and enhancement of theories and methods in the field of nondestructive testing. We demonstrate a simple technique for acquiring data from an impact test using a basic electronic stethoscope. The continuous wavelet transform (CWT) is used to process and display the transient responses of a bolted joint in a structure with bolts installed in a properly tightened manner and bolts incorrectly installed. Results for tight and loose bolts are compared using the semblance analysis approach described by Cooper and Cowan ["Comparing time series using wavelet-based semblance analysis," Computers & Geosciences 34, 95–102 (2008)]. We enhance the interpretation of results using speaker-driven audio output synthesized from CWT and semblance analysis.

10:45

3aSA9. Selective excitation of microcantilever array using ultrasound radiation force. Thomas M. Huber, Daniel C. Mellema, and Brad Abell (Dept. of Phys., Gustavus Adolphus College, 800 College Ave., St. Peter, MN 56082)

The symmetric and antisymmetric eigenstates of a coupled pair of 500 956 m length microcantilevers were excited using the ultrasound radiation force. The excitation was produced using the difference frequency between the two sidebands of a double sideband suppressed carrier AM (DSB-SC-AM) waveform centered on 500 kHz that was emitted by a pair of focused ultrasound transducers. A laser Doppler vibrometer measured the frequency response and deflection shapes of the cantilever pair. When the waveforms sent to the transducer resulted in radiation force from both transducers with the same phase, it excited the 10.00 kHz symmetric state of the cantilever array while suppressing the 10.17 kHz antisymmetric state. Similarly, when the radiation force from the two transducers was 180 degrees out of phase, it selectively excited the antisymmetric state while suppressing the symmetric state. This ability to selectively excite different vibrational eigenstates is a unique capability of this noncontact for modal excitation.

11:00

3aSA10. Limiting performance of active vibration compensation in free multi-degree-of-freedom mechanical systems. Vyacheslav Ryaboy (Newport Corp., 1791 Deere Ave., Irvine, CA 92606, vyacheslav.ryaboy @newport.com)

Examples of free (not supported) mechanical system requiring active vibration control range from space structures to vibration-isolated platforms for sensitive equipment. While active damping systems reducing resonance vibrations of the structure received significant attention and reached practical implementation, active vibration compensation, or cancellation of forced vibration, is still problematic in complex systems. In this work, active vibration compensation is considered in frame of limiting performance approach that aims at estimating the best results attainable in the whole class of linear multi-degree-of-freedom systems. Several design goals are considered, including absolute and relative displacements of various points of the system. The resulting estimates are valid for all possible sets of internal ac-

tive forces. The limiting performance is estimated in terms of frequency domain and work performed by active forces on system displacements, which is considered a measure of the control effort. These estimates are expressed as exact inequalities in explicit analytical form. Active force distributions emerging from limiting performance analyses can assist in finding rational system configurations and optimal controls. Minimum active work, expressed in terms of passive subsystem parameters, provide a criterion for optimization of a structure as a potential host for an active vibration control system.

11:15

3aSA11. Response of a simple system made complex by attaching subsystems—an analytic study. John J. McCoy (School of Eng., The Catholic Univ. of America, Washington, DC 20064)

A deceptively simple dynamical system comprised of a grounded mass to which are attached *N*, a large number, sprung masses has received much attention. The reported studies mostly involve physical experiments or numerical simulations. Given the large number of physical parameters needed to describe any single system, such a research agenda requires one to address an excessive number of experiment scenarios. Thus, the motivation for the analytical study reported. The focus of the study is the net force the attached subsystems cause to act on the base system. This force, while "internal" to the coupled system, is a response measure of primary importance since it accomplishes the work that determines the energy flow to and from the base system. As demonstrated, a further reason for the focus is an ability to construct an asymptotically valid, analytic expression for this response measure. Further demonstrated, using the constructed expression, is a resolution of a complex "process," that describing the energy flow to the system, by the action of an external forcing, and through the system, as a combination of simple "sub-processes." The resolution requires describing the complex process in a time/frequency phase space; the sub-processes are "local" in the phase space.

11:30

3aSA12. Wave vector method for estimating acoustic vector quantities. Derek C. Thomas and Kent L. Gee (Dept. of Phys. and Astronomy, Brigham Young Univ., N283 ESC, Provo, UT 84602)

Acoustic vector quantities such as particle velocity and intensity are usually estimated using the finite-difference technique. We present an alternate method, termed the wave vector method, for estimating vector quantities in acoustic fields in which the reactive intensity is small relative to the active intensity. This method utilizes the same equipment necessary for the finite-difference method, yet reduces sensitivity to sensor phase mismatch at low frequencies and eliminates bias error at high frequencies. The method is developed for arbitrary sensor configurations, and numerical results are presented for the special case of a regular tetrahedral configuration. [Work supported by NASA Stennis Space Center and STI Technologies.]

11:45

3aSA13. Impact of spherical probe scattering on estimation of acoustic vector quantities. Curtis P. Wiederhold (Dept. of Mech. Eng., Brigham Young Univ., 435 Crabtree Bldg., Provo, UT 84602, curtis.wiederhold@gmail.com), Kent L. Gee, Derek C. Thomas, Scott D. Sommerfeldt (Brigham Young Univ., Provo, UT 84602), and Jonathan D. Blotter (Brigham Young Univ., Provo, UT 84602)

Multimicrophone probes are often used to measure energy-based acoustical quantities. In some cases, these probes consist of microphones mounted on the surface of a sphere, which, due to the high level of symmetry, permits scattering effects to be better characterized. Scattering of a plane wave incident on a rigid sphere has been modeled to observe how particle velocity and intensity calculations are affected by the presence of the sphere. These effects have been investigated for the traditional "finite-difference" method and a recently developed "wave vector" estimation method. In the computer model, 3-D surface plots were made showing the calculated error as a function of angle of the incident plane wave and frequency. It is shown that using the wave vector method and purposefully orienting the probe in the planar sound field generally results in the most accurate measurements. [Work supported by NASA Stennis Space Center and STI Technologies.]

WEDNESDAY MORNING, 20 MAY 2009 GRAND BALLROOM I, 8:00 TO 11:50 A.M.

Session 3aSC

Speech Communication, Musical Acoustics, and Animal Bioacoustics: Source/Filter Interaction in Biological Sound Production (Lecture/Poster Session)

Ingo R. Titze, Chair
Dept. of Speech Pathology and Audiology, Univ. of Iowa, Iowa City, IA 52242-1012

Chair's Introduction—8:00

Invited Papers

8:05

3aSC1. Modulation of birdsong by resonance filters in the suprasyringeal vocal tract. Roderick A. Suthers (Medical Sci. and Dept. Biology, Indiana Univ., 1001 E. Third St., Bloomington, IN 47405, suthers@indiana.edu)

The sound generated in the avian vocal organ, the syrinx, is modified by the filter properties of the upper vocal tract before it radiates from the beak as song. Previous investigators have shown that the properties of this vocal tract filter are affected by changes in beak gape, but the basis of this effect remains controversial. In the present study, x-ray cinematography of singing birds was used to investigate the filter properties of the suprasyringeal vocal tract. These experiments show that the dimensions of the oropharynx and esophagus are actively controlled during song by motor patterns involving muscles of the hyoid apparatus. These song-related motor patterns, together with movements of the tongue, which modulate the opening from the mouth into the beak, adjust in real time the volume and

dimensions of the oropharynx and cervical esophagus so that the primary acoustic resonance of the upper vocal tract supports and tracks the fundamental frequency that is being generated by the syrinx. In juvenile Northern Cardinals, these song-related movements of the upper vocal tract first appear during subsong, while the beak is closed, and precede the changes in beak gape that typically accompany adult song. [Work supported by NIH.]

8:35

3aSC2. Source-filter interactions in birds—Theory and experimental evidence. Gabriel Mindlin, Ezequiel Arneodo (Dept. of Phys., Univ. of Buenos Aires, gabo.mindlin@gmail.com), and Franz Goller (Univ. of Utah, Salt Lake City, UT, goller@biology.utah.edu)

The diverse acoustic behavior of birds presents a rich source of natural vocalizations, some of which most likely reflect source-filter interaction. However, very little experimental evidence exists for the specific role of such interactions in shaping acoustic behavior or the functional relevance of the resulting acoustic features. As an example of how theoretical approaches can help explain acoustic observations, we discuss coexisting limit cycles in relaxation oscillators subjected to delayed feedback. The dynamical solutions of a nonlinear relaxation oscillator subjected to a delayed feedback are analyzed, and the equations under study are designed to model some aspects of the source-tract interaction in birdsong production. By deriving a phase equation for the system and analyzing its solutions, we are able to unveil analytical relationships between the parameters that lead to a variety of solutions. In particular, we study the coexistence of periodic solutions for similar pitch and first formant frequencies. These dynamical situations allow us to interpret the frequency jumps present in some birdsong syllables as an effect of the interaction between the source and the tract.

9:05

3aSC3. The epilaryngeal tube and air sacs in mammals as acoustic couplers—A review. Tobias Riede (Dept. of Biology, Univ. of Utah, 257 S 1400 E, Salt Lake City, UT 84112, t.riede@utah.edu) and Isao Tokuda (Japan Adv. Inst. of Sci. and Technol., Ishikawa, Japan)

The intralaryngeal space between glottis and the cranial edge of the larynx is called the epilaryngeal tube (synonyms are vestibulum larynges or supraglottal intralaryngeal space). Research on larynx models and in humans indicate that this part of the vocal tract filter can contribute to two important acoustic effects: (a) the singer's formant and (b) the enhancement of nonlinear source-filter coupling. The acoustic role of this structure in nonhuman mammals is unknown. Furthermore, in many species tube and sac-like cavities are attached to the epilaryngeal tube (laryngeal air sacs) and their function is equally little understood. In this presentation I will report on results testing the hypothesis that the shape (length and diameter) of the epilaryngeal tube in nonhuman mammals is related to morphological parameters (vocal fold morphology, larynx size, body size) and parameters characterizing the species vocal repertoire (repertoire size, maximum sound amplitude, fundamental frequency range, occurrence of nonlinear phenomena). Preliminary results indicate that the length of the epilaryngeal tube is a poor predictor of repertoire characteristics such as maximum sound amplitude and fundamental frequency range. However, species with a prominent epilaryngeal tube produce a large proportion of high fundamental frequency call types.

9:35

3aSC4. A possible role of nonlinear source-filter interaction in simulation of childlike speech. Brad H. Story (Dept. Speech, Lang., Hearing Sci., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721)

Simulation (or synthesis) of natural-sounding childlike speech has long been a challenge. This is likely due, at least in part, to an incomplete understanding of the nonlinear interaction of the voice source and the vocal tract filter. Speech production by children is typically characterized by a fairly high fundamental frequency of phonation and a short vocal tract length that produces high formant frequencies. Together, these two characteristics suggest that low-numbered harmonics (including the fundamental frequency) may often, or even necessarily, be in close proximity to one or more of the formant frequencies. Such conditions may lead to a strong interaction of the acoustic pressures in the vocal tract and the glottal airflow, and possibly the vibration of the vocal folds. The purpose of this study was to use kinematic models of the vocal folds and vocal tract shape, scaled to approximately represent a 5-year-old child, to generate individual vowels and sentences. Waveshape and harmonic content of glottal area, glottal flow, and radiated acoustic pressure signals were compared across three vowels and one sentence. [Research supported by NIH R01-DC04789.]

10:05—10:20 Break

10:20

3aSC5. Nonlinear source-tract interaction in singing. Ingo R. Titze (Dept. of Commun. Sci. and Disord., Univ. of Iowa, WJSHC, Iowa City, IA 52242 ingo-titze@uiowa.edu)

Most approaches to teaching singing are based on concept that there preferred vowel configurations for a given pitch. Speaking vowels are modified not only to create a variety of timbres, but also to support the sound source in self-sustained oscillation. Linear filter theory, traditionally applied to cans or strengthening by vocal tract coupling, or can it account for source bifurcations in vocal fold oscillation related to vowel selection. Singing are contrasted by selection of two fundamentally different mouth configurations, the megaphone shape and the inverted megaphone shape. With the megaphone shape (used by musical theatre and jazz), the second harmonic of the sound source is reinforced at all high pitches by keeping the first formant frequency high. With the inverted megaphone shape (used by classical singers of opera and art song), the second and third harmonics are reinforced by the second formant. A "lift" of the second harmonic over the first formant becomes a critical skill in acquiring a smooth timbre over a wide pitch range. [Work supported by NIH.]

All posters will be on display from 10:50 to 11:50. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 10:50 a.m. to 11:20 a.m. and contributors of even-numbered papers will be at their posters from 11:20 a.m. to 11:50 a.m.

3aSC6. Subglottal effects on the vowels across language: Preliminary study on Korean. Youngsook Jung (Speech Commun. Group, MIT, Cambridge, MA 02139, and Harvard-MIT Div. of Health Sci. and Technol., Cambridge, MA 02139)

Previous research suggested that subglottal resonances define vowel features [back] and [low] for English. In testing this hypothesis, we explored whether these vowel feature boundaries are independent of language. As preliminary research, we made recordings of speech and subglottal signals simultaneously for several adult Korean speakers. We found acoustic irregularities in F1 and F2 near subglottal resonances in Korean vowels. The boundary between [+low] and [−low] vowels agrees with speakers' first subglottal resonances (SubF1), while the boundary between [+back] and [−back] agrees with SubF2. In addition, we tested which of these hypotheses is correct for a low vowel in Korean, which has only one low vowel with no contrast of [+back] and [−back]: (1) the low vowel is always [+back] or (2) SubF2 is avoided but the low vowel is front or back depending on adjacent consonants. The measurements of F2 and F2sub were obtained in the context /CaC/, where C is a consonant. We found that SubF2 was always avoided for the low vowel. If the adjacent consonants were labial or velar, F2 of the low vowel was below SubF2, whereas if the consonants were alveolar, F2 of the vowel was above SubF2. [Supported by NIH /NIDCD Grant No. #DC00075.]

3aSC7. Automatic classification of consonant-vowel transitions based on subglottal resonances and second formant frequencies. Nancy F. Chen and Steven M. Lulich (Speech Commun. Group, MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

A growing body of work indicates that subglottal resonances (SGRs) divide the frequency space of vowels and consonants into distinct regions in several languages [S. Lulich, J. Phonetics (in press), and references therein]. For instance, the second formant frequency (F2) of back vowels lies between the first and second subglottal resonances (Sg1 and Sg2), while for front vowels F2 lies at higher frequencies than Sg2. In alveolar stop bursts, F2 lies between Sg2 and the third subglottal resonance (Sg3). In this study, we build on previous work [S. Lulich, JASA **124**, 2558 (2008)] to automatically determine SGRs using locus equations computed from 405 consonant-vowel (CV) transitions produced by an adult male native speaker of American English. Consonants and vowels are then classified into place of articulation and front/back features based only on the relations between F2 and the SGRs. The mean hit rate across categories was 77%, while the mean false alarm rate was 12%. These results show relative improvements of miss and false alarm rates by 21% and 14% from a baseline experiment, where F2 without SGRs was used to perform the classification. Our knowledge-based approach of classifying CV transitions is potentially useful when vowel and consonant identities are unavailable.

3aSC8. Source-filter interaction in the opposite direction: Subglottal coupling and the influence of vocal fold mechanics on vowel spectra during the closed phase. Steven M. Lulich (Speech Commun. Group, MIT, 77 Massachusetts Ave., Cambridge, MA 02139, lulich@speech.mit.edu), Matias Zanartu (Purdue Univ., West Lafayette, IN 47907), Daryush D. Mehta (Harvard-MIT Div. of Health Sci. and Technol., Cambridge, MA 02139), and Robert E. Hillman (Harvard Med. School, Boston, MA 02114)

Studies of speech source-filter interaction usually investigate the effect of the speech transfer function (loading) on vocal fold vibration and the voice source. In this study we explore how vocal fold mechanics affect the transfer function throughout the glottal cycle, with emphasis on the closed phase. Coupling between the subglottal and supraglottal airways is modulated by the laryngeal impedance. Although coupling is generally thought to occur only during the open phase of vocal fold vibration, a posterior glottal opening and the vocal fold tissue itself can allow sound transmission, thereby introducing coupling during the closed phase as well. The impedance of the vocal fold tissue at closure is shown to be small enough to permit coupling throughout the phonatory cycle, even in the absence of a posterior glottal opening. Open- and closed-phase coupling is characterized using mathematical models of the subglottal and supraglottal airways, and the parallel laryngeal impedances of the membranous glottis, posterior glottal opening, and vocal fold tissue. Examples from sustained vowels are presented, using synchronous recordings of neck skin acceleration, laryngeal high-speed videoendoscopy, electroglottography, and radiated acoustic pressure.

3aSC9. Nonlinear flow-structure coupling in a mechanical model of the vocal folds and the subglottal system. Michael S. Howe (College of Eng., Boston Univ., 110 Cummington St., Boston, MA 02215) and Richard S. McGowan (CReSS LLC, Lexington, MA 02420)

An analysis is made of the nonlinear interaction between flow in a mechanical model of the subglottal vocal tract and a model of the vocal folds. The mean flow through the system is produced by a "lung cavity" that is assumed to be steadily contracting. The lungs are connected to a subglottal tube of length L. The model for the vocal folds is at the other end of the subglottal tube. This model is a simple, self-exciting single-mass mathematical model of the vocal folds used to investigate the sound generated within the subglottal domain and the unsteady volume flux from the glottis. In the case when the absorption of sound by the lungs is presumed small and when the subglottal tube behaves as an open ended resonator (when L is as large as half the acoustic wavelength) a mild increase in volume flux is predicted. However, the strong appearance of second harmonics of the acoustic field is predicted at intermediate lengths, when L is roughly a quarter of the acoustic wavelength. In the cases of large lung damping, however, only modest changes in volume flux are predicted to appear. [Work supported by DC-009229 to UCLA under subcontract.]

3aSC10. An impedance-based inverse filtering scheme with glottal coupling. Matias Zanartu (School of Elect. and Comp. Eng., Purdue Univ., 206 S. Martin Jischke Dr., West Lafayette, IN 47907, mzanartu@purdue.edu), Julio C. Ho (Purdue Univ., West Lafayette, IN 47907), Daryush D. Mehta (Harvard-MIT Div. of Health Sci. and Technol., Cambridge, MA 02139), Robert E. Hillman (Massachusetts General Hospital, Boston, MA 02114), and George R. Wodicka (Purdue Univ., West Lafayette, IN 47907)

Source-filter interactions in voice production arise from considering finite glottal impedances that are comparable with that of the vocal tract and/or the subglottal system. Besides the changes in the source behavior, finite glottal impedances also allow for acoustic coupling between the subglottal and supraglottal tracts. Vowel dependence on the subglottal acoustic pressure is evident in recordings of skin acceleration overlying the suprasternal notch. To obtain estimates of other physiologic signals from skin acceleration, an impedance-based inverse filtering scheme that considers acoustic coupling is proposed. The scheme is based on an electrical analog model of a dipole source and assembles multiple approaches to estimate the system impedances. An experimental platform considers simultaneous measurements of skin acceleration, flexible high-speed videoendoscopy, oral volume velocity, intraoral pressure, electroglottography, and radiated acoustic pressure. Recordings of sustained vowels, co-articulated gestures, and pitch glides provide a comprehensive data set to study tract coupling and source-filter interactions. Preliminary results provide insights into the interrelation among the measured signals and illustrate the tissue instabilities of the vocal folds during some of the gestures. Further evaluation in cases of vocal fold pathology is also planned, where source-filter interactions are expected to be even more pronounced.

3aSC11. Vocal inharmonicity analysis: A promising approach for acoustic screening for dysphonia. Samuel Matteson (Dept. of Phys., Univ. of North Texas, 1155 Union Circle 311427, Denton, TX 76203-5017, Samuel.Matteson@unt.edu) and Fang-Ling Lu (Univ. of North Texas, Denton, TX 76203-5017)

In the United States, 6.8% of men, women, and children report voice problems currently, while during their lifetime ca. 29% will complain of some dysphonia, often due to pathologies of the vocal folds. The authors (a physicist and a speech pathologist) describe an interdisciplinary approach that exhibits promise for detecting physiological abnormalities of the vocal folds from a spectral analysis of the overtone series of spoken tokens. This preliminary investigation shows that the normal human vocal fold emits overtones that are very nearly precise integral values of the fundamental. Physiological problems of the vocal folds, however, introduce mechanical nonlinearities that manifest themselves as frequency deviations from the ideal harmonic (that is, integral) values. The authors quantify this inharmonicity, describing and illustrating how one can efficiently obtain and analyze such data. They outline, as well, a proposed investigation to assess the clinical sensitivity and significance of the analysis discussed in this work in conjunction with laryngoscopic diagnosis of pathology.

3aSC12. Geometric dependence of phonation threshold pressure and phonation onset frequency. Zhaoyan Zhang (UCLA School of Medicine, 31-24 Rehab Ctr., 1000 Veteran Ave., Los Angeles, CA 90095-1794)

Dependence of phonation threshold pressure and phonation onset frequency on vocal fold geometry was numerically investigated in a continuum model of the vocal folds. The geometric parameters of interest include the medial surface thickness, vocal fold depths of the body, and cover layers. Changes in these parameters may occur either due to laryngeal muscle activation, vocal pathology, or aging. Simulations show that both threshold pressure and onset frequency increased with decreasing vocal fold depth. However, depending on glottal geometry, increasing medial surface thickness may either increase or decrease phonation threshold pressure and frequency. Variation of the medial surface thickness also led to phonation onset occurrence at a different eigenmode, leading to sudden changes in onset frequency and vocal fold vibration pattern. These numerical predictions were further supported by experiments using a self-oscillating physical model. The result of this study confirms previous findings that phonation onset occurs as a matching between the flow stiffness, which is proportional to subglottal pressure, and vocal fold stiffness, which scales with natural frequency squared and depends on geometry and boundary conditions of the vocal fold structure as well as material properties of the vocal folds. [Work supported by NIH.]

3aSC13. Analysis and synthesis of humpback whale songs. Pranab K. Dhar, Sangjin Cho, Jong-Myon Kim, and Uipil Chong (Univ. of Ulsan, Ulsan 680-749, Korea)

The humpback whale song is one of the most complex, non-human, acoustic displays in the animal kingdom. In this paper, we analyze and synthesize the humpback song using additive synthesis which is one of the most powerful techniques for the analysis, modification, and synthesis of complex audio or speech signals. This method creates complex sounds by adding together individual sinusoidal signals called "partials." A partial's frequency and amplitude are each time-varying functions, so it is a more flexible version of the harmonic associated with a Fourier series decomposition of a periodic waveform. Applying the fast Fourier transform to all overlapping frames (small time segments), we can detect peaks from the frequency spectrum of each frame and calculate the frequency, magnitude, and phase for all the peaks. These extracted parameters are then used as components of the sinusoidal signals (partials) for the synthesis process. This process finally reconstructs and concatenates each frame, generating the synthesized humpback song. We observed that this synthesized song resembles much more closely the original humpback song by the time/frequency domain representation and listening. In the future, we will analyze and synthesis other baleen whale songs with this powerful synthesis method. [Work supported by IITA (IITA-2008-(C1090-0801-0039)).]

WEDNESDAY MORNING, 20 MAY 2009 PARLOR B/C, 7:45 TO 11:30 A.M.

Session 3aSP

Signal Processing in Acoustics and Acoustical Oceanography: Poroelastic Materials: Models, Bounds, and Parameter Estimation

Max Deffenbaugh, Chair
ExxonMobil Research & Engineering Co., 1545 Route 22 East, Annandale, NJ 08801

Chair's Introduction—7:45

Invited Papers

7:50

3aSP1. Acoustics in granular porous media and the consolidation continuum. James G. Berryman (Earth Sci. Div. Lawrence Berkeley Natl. Lab, One Cyclotron, Dr. Berkeley, CA 94720, jgberryman@lbl.gov)

One key influence on acoustic wave propagation in granular porous media is the degree to which the solid constituent grains can be treated as fully consolidated, weakly and/or partially consolidated, or entirely unconsolidated. Well-known results of Gassmann and Biot all implicitly assume that the medium is fully consolidated, with welded contacts at the grain-to-grain interfaces. Such implicit assumptions are also present in standard bounding methods for elastic behavior, such as the Hashin-Shtrikman bounds of elasticity. Influence of reductions in the degree of consolidation can be quantified for partially or weakly consolidated media, and granular media having no welded contacts can also be treated as porous media for sound waves in the Biot-Gassmann sense when grains are experiencing sufficient overall confining stress. Methods recently developed to quantify wave propagation characteristics over the full range of possible behaviors in grain packs, as well as comparisons to experiments, will be presented.

8:20

3aSP2. Fabric dependence of poroelastic wave propagation. Stephen Cowin and Luis Cardoso (Dept. of Biomed. Eng., The City College, NY, NY 10031, cardoso@soe.ccny.cuny.edu)

The governing equations for wave motion in the linear theory of anisotropic poroelastic materials are developed and extended to include the dependence of the constitutive relations upon fabric. Fabric is a quantitative stereological measure of the degree of structural anisotropy in the pore architecture of a porous medium. With the addition of the second order symmetric tensor fabric variable, the formulation of wave motions in the poroelastic theory is consistent with the presentations of Biot and later authors. The dependence of all the material tensors in the Biot theory on the fabric tensor aligns the material tensors, so when a single direction that is a plane of material symmetry is selected, the equations simplify considerably. The customary polynomial of the sixth deg for the wave speeds of the anisotropic Biot theory in a selected direction reduces to three quadratic equations to solve for the wave speeds and attenuation. While the theory is applicable to any saturated porous material, the two longitudinal waves predicted by this model are measured in cancellous bone and used to derive the corresponding anisotropic elastic constants. Representative examples of bone loss are analyzed as a function of the porosity, tissue density, and fabric.

8:50

3aSP3. Anisotropic bone loss characterized using fabric-based poroelasticity. Luis Cardoso and Stephen Cowin (Dept. of Biomedical Eng., The City College, 138th St., and Convent Ave., NY, NY 10031, cardoso@engr.ccny.cuny.edu)

The current gold standard to assess bone loss is the bone mineral density (BMD) test, performed with a dual x-ray absorptiometry system. BMD measurements are highly correlated to bone mass when measured in spine, wrist, and femoral neck. However, the BMD cannot fully explain the decrease in bone strength and the associated risk of fracture because bone loss is not homogeneous, and does not depend on mass changes only. In this study, a fabric-based poroelastic model is used to determine the elastic constants of bone that have undergone disuse-osteoporosis and bone treated with an antiresorptive drug. Images from microCT are used to determine the porosity, density, and fabric on each group of specimens. These measurements are used in the proposed fabric-based poroelastic model to determine the sample's anisotropic elastic constants. These elastic constants are then compared to the ones obtained by ultrasound wave propagation measurement as previously described in the Cowin and Cardoso presentation. Image analysis and ultrasound wave propagation are both shown able to determine changes in anisotropic elastic constants during bone loss when used with the proposed anisotropic poroelastic model. This study provides the foundation to define a bone quality criterion beyond the sole determination of bone mass.

Contributed Paper

9:20

3aSP4. Effective medium models from x-ray microtomography images. Max Deffenbaugh, John Dunsmuir (ExxonMobil Res. and Eng. Co., 1545 Rte. 22 East, Annandale, NJ 08801, max.deffenbaugh@exxonmobil.com), Shiyu Xu (ExxonMobil Exploration Co., Houston, TX 77060), and Enru Liu (ExxonMobil Upstream Res. Co., Houston, TX 77252)

A novel method is described where an effective medium model is generated from a 3-D x-ray microtomography image of a rock sample. In contrast to current modeling practice, material phases are not assigned idealized geometries, like spheroids. Instead, strain concentration tensors are computed numerically for the true phase geometries observed in the 3-D image. In this method, a 3-D image of x-ray attenuation is converted into a 3-D elastic properties volume, the strain field is computed within the volume, and the strain concentration tensors, which quantify the contribution of each grain and pore type to the bulk elastic properties, are calculated from the strain field. From only one representative sample of a rock type and with no assumptions about grain or pore shape, an effective medium model is generated that specifies the effective elastic tensor as a function of the various mineral and pore volume fractions. The method is demonstrated by generating effective medium models for several rock types. These models can be used in hydrocarbon exploration to relate the velocities of seismic waves in a poroelastic reservoir to the porosity, structure, and composition of the rock matrix and the type and distribution of the fluid within the pore spaces.

Invited Papers

9:35

3aSP5. Model parameter estimation of a porous seabed using *in situ* acoustic-probe and chirp-sonar reflection data. Altan Turgut (Naval Res. Lab., Acoust. Div., Washington, DC 20375)

The velocity dispersion and frequency dependency of attenuation in marine sediments are two important measures that provide the ability to test the validity of geoacoustic models as well as to estimate physical properties from the acoustic data. A widely-used Biot–Stoll theory for elastic-wave propagation in porous marine sediments seems to predict a narrow band of dispersion that was measured only in well-sorted granular marine sediments [A. Turgut and T. Yamamoto, J. Acoust. Soc. Am. **87**, 2376–2382 (1990)]. Recent measurements in silty-sand sediments show almost linear frequency dependency of attenuation and mild velocity dispersion within the 3–80 kHz frequency band. The newly measured attenuation and velocity dispersion curves are in good agreement with those predicted by an extended Biot theory for sediments with a distribution of pore sizes [T. Yamamoto and A. Turgut, J. Acoust. Soc. Am. **83**, 1744–1751 (1988)]. Simultaneous measurements of *in situ* acoustic-probe data and chirp-sonar reflection data are also used to estimate the model parameters of the extended Biot model. [Work supported by the Office of Naval Research.]

10:05—10:15 Break

10:15

3aSP6. Cancellous bone as a poroelastic medium: Extracting underlying material properties from improved ultrasonic measurements of frequency dependent attenuation and phase velocity. James G. Miller, Christian C. Anderson, Adam Q. Bauer, Karen Marutyan, G. Larry Bretthorst, Mark R. Holland (Washington Univ., 1 Brookings Dr., St. Louis, MO 63130), Brent K. Hoffmeister (Rhodes College, Memphis, TN 38112), and Keith A. Wear (U.S. Food and Drug Administration, Silver Spring, MD 20993)

Cancellous bone can be modeled as a poroelastic medium consisting of a viscous fluid (marrow) and a stiff framework (trabeculae). Transmission of ultrasound through bone can result in fast and slow waves as well as mode conversion and multiple scattering, all of which can mask the underlying material properties of bone that determine fracture risk. Phase cancellation at the receiving transducer and interference effects in the field can further compromise the potential of bone sonometry for assessing bone quality. The attenuation coefficient is observed to increase approximately linearly with frequency, a fact that accounts for the use of broadband ultrasound attenuation (BUA) to characterize bone. Many laboratories report that the phase velocity of ultrasonic waves propagating through cancellous bone decreases with increasing frequency (negative dispersion). However, an increase in phase velocity with frequency (positive dispersion) would be expected based on the causality-imposed Kramers–Kronig relations. This apparent negative dispersion in bone can arise from interference between fast wave and slow wave longitudinal modes. Interference and phase cancellation at the receiver can also be sources of artifactual alterations in the attenuation and phase velocity. [Work supported in part by NIH R37HL40302 R01AR057433, NSF CBET-0717830].

Contributed Papers

10:45

3aSP7. Successful use of Bayesian inference to obtain the fast and slow wave properties of cancellous bone from sonometry measurements. Christian C. Anderson (Washington Univ., 1 Brookings Dr., St. Louis, MO 63130), Michal Pakula, Pascal Laugier (Univ. Paris 6, 75006 Paris, France), Adam Q. Bauer, Karen Marutyan, G. Larry Bretthorst, Mark R. Holland, and James G. Miller (Washington Univ., St. Louis, MO 63130)

Clinical trials have demonstrated that frequency dependent attenuation [broadband ultrasound attenuation, (BUA)] and speed of sound (SOS) provide an assessment of fracture risk. Previous studies from our laboratory of the "forward problem" show that the negative dispersion observed in cancellous bone can arise when signals consisting of overlapped fast and slow waves are analyzed as if only one wave is present. The present study examines the "inverse problem" using Bayesian probability methods suitable for recovering the underlying material properties from received transmission sonometry signals complicated by the presence of multiple signals. Transmission studies were carried out on bone mimicking phantoms and on specimens of femur condyles. Received rf signals served as input to our Bayesian algorithm. The program that implements the Bayesian calculations reconstructs the individual fast and slow waves and constructs the model output from parameters that maximize the joint posterior probability. Composite rf signals constructed by summing these waves were in excellent agreement with signals transmitted through bone-mimicking phantoms and were in good agreement with signals transmitted through bone samples despite the additional complexities arising from ultrasound propagation through the complex trabecular network. [Work supported in part by NIH R37HL40302 R01AR057433, and NSF CBET-0717830].

11:00

3aSP8. Geoacoustic inversion performed from two source-receive arrays in shallow-water waveguide. Alexey Sukhovich, Philippe Roux, and Marc Wathelet (LGIT, UMR CNRS 5559, Univ. Joseph Fourier, Grenoble, France, Alexey.Sukhovich@obs.ujf-grenoble.fr)

Raylike propagation of acoustic waves in a shallow-water waveguide between two vertical line arrays is investigated by applying a double beamforming algorithm, which performs time-delay beamforming on *both* emitting and receiving arrays and allows identification of eigenrays by their emission and reception angles and arrival times. From the intensity of each eigenray, it is possible to determine reflection coefficient from the bottom of the waveguide as a function of an angle of incidence. The procedure was initially tested in a small-scale tank experiment for an acoustic waveguide with either steel or Plexiglas bottom. By fitting an experimentally found reflection coefficient with a corresponding theoretical expression, an estimate for the speed of shear waves in the bottom material was obtained. Similar analysis was subsequently applied to the data obtained during an at-sea experiment, which was performed between two vertical transducer arrays in shallow-coastal waters of the Mediterranean. An angle-dependent bottom reflection coefficient was extracted and geoacoustic inversion was performed by fitting the data with theoretical calculations, in which bottom sediments were modeled as a multilayered system. Good agreement between experiment and theory was observed. Our results indicate possible application of eigenray intensity analysis based on double beamforming algorithm for geoacoustic inversion problems.

11:15

3aSP9. Monitoring stress related velocity variation in concrete with a 2×10^{-5} relative resolution using diffuse ultrasound. Larose Eric (Lab. de Geophys. interne et Tectonophysique, CNRS & Univ. J. Fourier, Grenoble, France) and Hall Stephen (CNRS & Grenoble Univ., Grenoble, France)

Ultrasonic waves propagating in solids have stress-dependent velocities. The relation between stress (or strain) and velocity forms the basis of nonlinear acoustics. In homogeneous solids, conventional time-of-flight techniques have measured this dependence with spectacular precision. In heterogeneous media such as concrete, the direct (ballistic) wave around 500 kHz is strongly attenuated and conventional techniques are less efficient. In this manuscript, the effect of weak stress changes on the late arrivals constituting the acoustic diffuse coda is tracked. A resolution of 2×10^{-5} in relative velocity change for stress change is attained. We are thus sensitive to stress change much lower than 50 kPa. This provides an original way to measure the nonlinear parameter with relatively small stress variations (less than 100 kPa).

Session 3aUW

Underwater Acoustics and Acoustical Oceanography: Physics-Based Undersea Clutter Model Verification and Validation I

Juan I. Arvelo, Cochair
Applied Physics Lab., Johns Hopkins Univ., Laurel, MD 20723-6099

Kenneth G. Foote, Cochair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Timothy K. Stanton, Cochair
Dept. of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Chair's Introduction—8:00

Invited Paper

8:05

3aUW1. Dynamic and acoustic modeling of discrete biologics for active sonar simulation. Brian R. La Cour and John E. Hamann (Appl. Res. Labs., The Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029)

A model is proposed for simulating the group dynamics and acoustic echoes of aggregate, discrete biologics in midfrequency (2–5 kHz) active undersea sonar. Using a Lagrangian description, group behavior is modeled by a set of Newtonian forces and stochastic perturbations on a collection of point masses. Interactions comprise a grouping force, which gives a characteristic separation distance, and an arrayal force, which tends to align the velocities of the group's constituents. The group dynamics is characterized in terms of their statistical and asymptotic properties and studied as a function of the model parameters. Several models for the acoustic response are considered and studied for both individual and aggregate response. Finally, the resulting synthetic echoes are compared against those of actual biologics measured on a midfrequency active sonar system. The synthetic echoes are found to reproduce well both the qualitative characteristics of measured biologics and the resulting tracking and classification features found in a typical active sonar signal and information processing chain.

Contributed Paper

8:25

3aUW2. Coherent backscattering effect from mid-frequency reverberation measurements in shallow-water. Karim G. Sabra (School of Mech. Eng., Georgia Inst. of Technol., 771 Ferst Dr. NW, Atlanta, GA 30332-0405)

The coherent backscattering (or backscattering enhancement) effect is a persistent interference effect which occurs between waves propagating in random media such as randomly distributed scatterers and rough surfaces. Previous experimental and theoretical studies in optics, acoustics, or seismology have demonstrated that the coherent backscattering results in a doubling of the recorded intensity reflected exactly in the backward direction with respect to any other direction. We will present experimental observations of the coherent backscattering effect from mid-frequency reverberation data (3–4 kHz) collected in shallow water off the west coast of Italy using various sensor array configurations. This demonstrates the presence of coherent effects in shallow water reverberation recordings from rough seabeds. Hence, coherent reverberation effects need to be accounted for to obtain accurate predictions of sea clutter levels. Furthermore, accurate rendering of the coherent backscattering effect could be used to simply verify and validate physics-based undersea clutter models. Practical implementation for the design of active sonar systems will be discussed. [Work supported by ONR, Code 321.]

Invited Papers

8:40

3aUW3. Measurements and modeling of midfrequency clutter from fish aggregations over Georges Bank in the Gulf of Maine. Roger C. Gauss, Joseph M. Fialkowski, Richard Menis, Edward L. Kunz (Naval Res. Lab., Code 7140, Washington DC 20375-5350, roger.gauss@nrl.navy.mil), Timothy K. Stanton, Cynthia J. Sellers (Woods Hole Ocean. Inst., Woods Hole, MA 02543), and J. Michael Jech (NOAA NE Fish. Sci. Ctr., Woods Hole, MA 02543)

Broadband midfrequency acoustic measurements of fish backscattering were made with two complementary sonar systems in September 2008 at the northern slopes of Georges Bank in the Gulf of Maine. One system, towed near the surface, was downward-looking (short-range: < 0.2 km) and was used to image fish at high vertical resolution throughout the water column. The other system was horizontal-looking (relatively long-range: 1–10 km), consisting of a vertical source array and a horizontal receive array, and was used to image fish at high horizontal resolution. In this talk, the focus is on characterizing the longer-range data. Spectrogram analysis showed that the echoes had resonance frequencies in the 2–4 kHz band, consistent with scattering by Atlantic herring, while examination of normalized match-filter output revealed strong, spatiotemporally variable clutter. The associated probability-density functions

(PDFs) exhibited variable non-Rayleigh behavior. Several PDF models were fit to the data, with the K and Poisson-Rayleigh PDF models providing not only good fits, but physical insights into the clutter process (such as estimates of the expected number of discrete scatterers per unit area). These results provide an improved understanding of both fish behavior and the characteristics of fish aggregations as long-range clutter fields. [Work supported by ONR.]

9:00

3aUW4. Bioalpha, transmission loss and validation of scattering models in biologically intense shallow water environments. Orest Diachok (Appl. Phys. Lab. Johns Hopkins Univ., Laurel, MD 20723-6099)

Scattering and extinction by fish swim bladders are manifestations of the same phenomenon. Scattering causes reverberation from dispersed fish and distinct echoes from fish schools (clutter), and diminishes the magnitude of transmitted signals. The biological extinction coefficient (bioalpha) has been shown to have a large effect on transmission loss (TL) in several biologically intense, shallow water environments, including the Bristol Channel [Weston (1967)], Yellow Sea [Qiu *et al.* (1999)], Gulf of Lion [Diachok (1999)], and Santa Barbara Channel [Diachok (2005)]. Consequently, models of biological scattering must include the effects of bioalpha on TL in shallow water environments. The objectives of this paper are to review the experimental evidence of the effects of bioalpha on TL, and to describe and demonstrate the effectiveness of an analytical approach to modeling this phenomenon. Since it is not possible at present to predict the effects of bioalpha on TL from first principles (distributions of swim bladder dimensions, spatial distributions of number densities, and timing of diurnal changes in school structures and seasonal migration patterns are only approximately predictable), experiments designed to demonstrate the validity of scattering models in biologically intense, shallow water environments should include concurrent broadband TL measurements during both night and day.

Contributed Paper

9:20

3aUW5. Statistics of echoes from a directional sonar beam insonifying finite numbers of single scatterers and patches of scatterers. Dezhang Chu (NOAA/NMFS/Northwest Fisheries Sci. Ctr., Seattle, WA 98112) and Timothy K. Stanton (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

When a sonar beam sweeps across a field of scatterers, the echoes can fluctuate significantly from ping to ping. The fluctuations can be strongly non-Rayleigh because: (1) there can be a small number of scatterers in the beam; (2) the distribution of scatterers can be inhomogeneous or "patchy;" and (3) the echoes are weighted by the nonuniform response of the sonar beam. In this paper, a general formulation combining equations derived by Ehrenberg [Proc. Conf. Eng. Ocean Environ. 1, 61–64, (1972)] and Barakat [Optica Acta 21, 903–921, (1974)] is developed to account for a directional sonar beam involving an arbitrary finite number of scatterers, each with an arbitrary echo probably density function (PDF) and randomly located in the beam. Theoretical predictions are made, along with numerical simulations for validation, for a range of conditions, including: (1) different number of scatterers arbitrarily located in the beam and (2) different echo PDFs of the scatterers. Here, a "scatterer" could be an individual or a patch of scatterers whose dimensions are much smaller than the footprint of the sonar beam. Although the application is intended for volumetric patches, the formulation could be applied to areal patches under appropriate conditions. [Work supported by ONR.]

Invited Papers

9:35

3aUW6. Comprehensive, integrated database of elements contributing to bottom and volume clutter. Kathleen J. Vigness-Raposa, William T. Ellison, Adam S. Frankel, David Zeddies, Cheryl L. Schroeder (Marine Acoust., Inc., 809 Aquidneck Ave., Middletown, RI 02842, kathleen.vigness@marineacoustics.com), Peter V. August, Christopher Damon, and Charles LaBash (Univ. of Rhode Island, Kingston, RI 02881)

Multiple components within and at the bottom of the water column can scatter acoustic signals, resulting in decreased sonar performance and compromised tactical decision making. Many of the biological organisms that contribute to acoustic clutter are also protected or managed species, integral to a fully functional marine ecosystem. In order to understand, predict, and model bottom and volume scattering, detailed descriptions of the spatial, spectral, and temporal variability of clutter elements are needed, along with reliable estimates of backscattering strength. All of these necessary components are dependent on marine wildlife databases. Efforts by several organizations have resulted in databases that provide pieces of the overall puzzle. Databases of distribution and abundance (Census of Marine Life's OBIS-SEAMAP), two-dimensional movement (Census of Marine Life's Tagging of Pacific Pelagics), and diving and movement behavior (Marine Wildlife Behavior Database) are examples of current, but disparate, projects. Ongoing efforts are consolidating inputs from such databases to predict realistic, spatially dependent bioclutter animats. By linking geospatial features of clutter elements in training and mission planning systems, sonar operators and designers can better predict and analyze the occurrence of target-like clutter. [Work supported by NAVSEA SBIR, ONR STTR, and NOAA.]

9:55—10:15 Break

10:15

3aUW7. Modeling persistent clutter. Bruce K. Newhall (Johns Hopkins Univ., Appl. Phys. Lab, 11100 Johns Hopkins Rd., Laurel, MD 20723)

Clutter is the collection of targetlike returns from nontarget sources that are observed in active sonar use. Persistent clutter is the set of clutter returns that persist across multiple transmissions and thereby form false target tracks in sonar processing. The U.S. Navy is developing active sonar simulators for training purposes. Effective training requires physically realistic clutter simulation including modeling of persistent clutter. Potential physical sources of persistent clutter are discussed, and a variety of modeling approaches are reviewed. Measures of effectiveness for model comparison with persistent clutter data are given.

10:35

3aUW8. Exploring spatiotemporal distribution as a clutter simulation validation measure. Juan I. Arvelo, Jr. (The Johns Hopkins Univ., Appl. Phys. Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099)

The amplitude distribution is widely used to compare modeled and measured undersea clutter [J. Arvelo *et al.*, JASA **120**, 3381 (2006)]. However, additional validation measures must be identified to ensure that clutter models are suited to realistically simulate active sonar displays and accurately emulate tracker performance. For example, the amplitude distribution ignores the stochastic arrangement of clutter at various space/time scales [V. Lupien, JASA **105**, 2187–2282 (1999)]. Active midfrequency vertical line array data from the ASIAEX 2001 experiment in the East China Sea is processed to explore the suitability of the clutter spatiotemporal distribution as a key validation measure. The vertical array permitted the separation of bottom from surface generated clutter. A physics-based seafloor clutter model [B. Newhall and J. Arvelo, JASA **118**, 2041 (2005)] was implemented to simulate the experimental scenario with known measured and inferred environmental conditions. Comparisons of the measured and simulated amplitude and spatiotemporal distributions uncovered significant effects that must be included to improve existing clutter models. [This effort was conducted under the auspices of the Undersea Warfare Business Area Independent Research and Development program of the Johns Hopkins University Applied Physics Laboratory.]

10:55

3aUW9. Midfrequency active sonar clutter statistics segregated by broad clutter types. James M. Gelb (Appl. Res. Labs., UT Austin, 10000 Burnet Rd., Austin, TX 78758, gelb@arlut.utexas.edu)

The empirical distributions of normalized matched-filter echoes from a midfrequency active sonar with hyperbolic frequency-modulated waveforms in a myriad of oceanic environments are studied for three broad clutter classes: bottom structures, diffuse compact clutter (e.g., seaweed), and compact nonstationary clutter (e.g., biologics). The distributions are characterized using the K distribution and the generalized Pareto distribution. Methods of parameter estimation are discussed, and parameters are computed for small subregions of the clutter fields. A plot of the Kolmogorov–Smirnov goodness-of-fit statistic of individual subregions is presented for each model and class to highlight the versatility of the models when applied to large quantities of data. Cumulants are computed from the data and are utilized as features in a classifier to demonstrate separability between the classes. An important aspect of this work is the use of distinct clutter classes as opposed to collectively characterizing all clutter as reverberation. Environmental effects are not considered, as the goal of this work is to determine the utility of local clutter estimation models in practical sonar processing systems where accurate environmental data is unavailable. [This work is sponsored by the Office of Naval Research undersea signal processing discovery and invention program.]

11:15

3aUW10. Simulation of rippled-sand sonar imagery with visual and statistical verification. Shawn F. Johnson (Graduate Prog. in Acoust., Appl. Res. Lab., 117 Appl. Sci. Bldg., State College, PA 16804)

Synthetic aperture sonar (SAS) imagery is often characterized by a decidedly non-Rayleigh pixel amplitude distribution, owing to its inherent high-resolution combined with speckle induced by the coherent image formation process. Recent work has shown increasing the resolution cell size by limiting the image bandwidth typically produces images with statistics tending toward a Rayleigh distribution. A caveat to this generalization is that the trend toward Rayleigh is dependent on orientation of the SAS system to strongly correlated sea-floor structures such as sand-ripples. A method has been developed to simulate rippled-sand sonar imagery at high-frequencies (i.e., on the order of 100 kHz), which accounts for nonsymmetric ripple shape, sediment acoustic properties, sonar to ripple orientation, system resolution, and coherent imaging induced speckle. This numerical simulation method is computationally inexpensive and compares well both visually and statistically with collected data over a wide range of orientation angles. Simulation methods will be presented with verification using SAS imagery collected by the Naval Surface Warfare Center: Panama City Division. [Work supported by ONR Grants No. N00014-04-1-0013 and No. N00014-06-1-0245, and Code 32.]

11:35

3aUW11. Experimental tests of models for high-frequency scattered envelope statistics. Anthony Lyons (Appl. Res. Lab., Penn State Univ., State College, PA 16804, apl2@psu.edu)

There has been, in general, a lack of quantitative comparison between high-frequency physics-based clutter model predictions and experimental data due mainly to difficulties in performing well-controlled at-sea experiments and obtaining necessary ground truth. To address this shortfall, experiments have been performed over the past 5 yrs to gather data with which to confront recently developed physics-based clutter models. The experimental data sets were obtained during experiments conducted off of Elba Island, Italy in May 2003 and at Seneca Lake, New York, in the summers of 2005 and 2006. Colocated high-frequency, broadband acoustic data and high-resolution environmental ground truth data were collected as part of these experiments. In all cases comparisons between model predictions and measured data showed promising similarities in *K*-distribution shape parameter estimates and their dependence on bandwidth, grazing angle, and seafloor type. The favorable model/data comparisons show that it is possible to link the scattered envelope distribution to measurable environmental properties, providing the foundation necessary for solving several important problems related to the detection of targets in non-Rayleigh clutter including performance prediction for different systems based on seafloor properties and extrapolation of performance to other systems and bandwidths.

Session 3pAAa

Architectural Acoustics: Absorption Testing and Characterization

Ron Sauro, Chair
NWAA Laboratories Inc., 25132 Rye Canyon Loop, Santa Clarita, CA 91355

Contributed Papers

1:00

3pAAa1. Impedance tube absorption testing of poly blend materials presently used in the carpet industry. Ruth Anne Mazur (Columbia College Chicago, 237 E. Wildflower Ln., Round Lake Beach, IL 60073, ruthie.mazur@sbcglobal.net)

Cyril M. Harris authored a paper in 1955 titled "Acoustical Properties of Carpet" in which absorption coefficients for carpet were examined according to specific parameters, including fiber content. Since then, new fibers and blends have been introduced into the carpet industry. This study aims to compare absorption coefficients found from testing of these newer fibers. New blends that are now commonly used and sold in the carpet industry include polyester, polypropylene, and polyurethane. Using a Bruel & Kjaer Two Microphone Impedance Measurement Tube and Bruel & Kjaer PULSE LabShop, normal incidence absorption coefficients are gathered in order to compare the coefficient curves for these materials. Testing is performed in accordance with the ASTM E 1050-98 standard. Preliminary testing on carpet with these fibers shows lower values from 3.5 to 6.5 kHz in the curves for polypropylene carpet with padding compared to other fibers with padding. Several parameters of the preliminary samples were not constant and may have affected the results. Ongoing analysis and testing of newer fibers will enable comparison with previous studies. Absorption coefficients and parameters for carpet available in acoustical modeling programs will also be assessed in relation to the results. [Work supported by Columbia College Chicago.]

1:15

3pAAa2. Felt panels for an art space. Jessamyn A. Newcomb (Dept. of Architecture, Virginia Polytechnic and State Univ., 201 Cowgill Hall (0205), Blacksburg, VA 24061)

The art building on Virginia Tech's campus is an adaptive reuse project. Despite its current function, the transformation of the space has left it acoustically ill suited for class instruction. This research focuses on absorption of specific frequencies through the use of felt panels to reduce noise and increase speech intelligibility for class instruction. Microperforated panels are proposed as a solution due to their simple structures and predictable wideband absorption characteristics. Benefits will be quantified by measuring pink noise decay in the building pre- and postpanel installation.

1:30

3pAAa3. Absorption coefficients part 1: Is square area enough? Ron Sauro, Michael Vargas (NWAA Labs., Inc., 25132 Rye Canyon Loop, Santa Clarita, CA 91355, audio_ron@msn.com), and Gary Mange (Western Electroacoustics Labs., Santa Clarita, CA 91355)

Attempting to correlate the measurements of absorption coefficients in ASTM-C423, ISO-354, and ISO-17497-1, it was observed that the results were varying to a large degree. An experiment was set up with 1 in., 6 lb density fiberglass panels having different shapes, sizes, and parameters. The shapes, sizes, and parameters of the panels are described in this paper. The types of experiments are also described along with the process that was used based on input from many of the members of the industry. The end results and conclusions are described in a paper titled "Absorption coefficients part 2: Is "edge effect" more important than expected?"

1:45

3pAAa4. Absorption coefficients part 2: Is "edge effect" more important than expected? Ron Sauro, Michael Vargas (NWAA Labs., Inc., 25132 Rye Canyon Loop, Santa Clarita, CA 91355, audio_ron@msn.com), and Gary Mange (Western Electro-Acoust. Labs., Santa Clarita, CA 91355)

Attempting to correlate the measurements of absorption coefficients in ASTM-C423, ISO-354, and ISO-17497-1, it was observed that the results were varying to a large degree. An experiment was set up as described in a paper titled "Absorption coefficients part I: Is square area enough?" It was observed that the edges of the sample had a significant effect on the absorption coefficient. The results of these experiments are presented in numerical and graphic form in this paper along with conclusions and suggested corrections to the "absorption coefficient" as used in all calculations in acoustics.

Session 3pAAb

Architectural Acoustics: American Institute of Architects Short Course

K. Anthony Hoover, Chair
McKay Conant Hoover, Inc, 5655 Lindero Canyon Rd., Suite 325, Westlake Village, CA 91362

Chair's Introduction—1:00

Invited Papers

1:05

3pAAb1. Introduction to workshop goals. K. Anthony Hoover (McKay Conant Hoover, Inc., 5655 Lindero Canyon Rd., Westlake Village, CA 91362) and Bennett M. Brooks (Brooks Acoust. Corp., Vernon, CT 06066)

Many states require that architects obtain continuing education unit (CEU) credits annually, in order to maintain their registration or licensure. The American Institute of Architects (AIA) Continuing Education System (CES) offers continuing education courses, which may be given by third party providers. The Technical Committee on Architectural Acoustics (TCAA) is an AIA/CES Registered Provider. The goal of this workshop is to prepare members of the TCAA so that they may be authorized to present a short course which can earn attendees CEU credit. In order for TCAA members to qualify to meet the AIA requirements, they must attend this workshop, which is given in two parts. The first part will be devoted to the short course presentation material, which is in a standardized format. The second part of the workshop will focus on the AIA CEU Program registration and reporting requirements. Of course, anyone is free to register with the AIA to provide their own CEU program. However, the advantages of participating in this program are that the TCAA short course is already prepared, is preapproved by the AIA, and the registration fees are paid by the Acoustical Society of America.

1:25

3pAAb2. Architectural acoustics: Short course presentation material. K. Anthony Hoover (McKay Conant Hoover, Inc., 5655 Lindero Canyon Rd., Westlake Village, CA 91362)

The Technical Committee on Architectural Acoustics (TCAA) is a Registered Provider in the American Institute of Architects (AIA) Continuing Education System (CES). The TCAA has developed a standardized introductory short course for architects, called "Architectural Acoustics." An architect can earn one continuing education unit (CEU) by attending this short course, if it is presented by a qualified member of TCAA. The course covers topics in sound isolation, mechanical system noise control, and finish treatments. This paper will cover the course material in order to prepare and qualify potential presenters. There will be time at the end of the paper for questions and answers on the course material. In order to qualify as an authorized presenter for this AIA/CES short course, attendance at this workshop and membership in TCAA are required.

2:25

3pAAb3. Provider registration and reporting requirements. Bennett Brooks (Brooks Acoust. Corp., 30 Lafayette Sq., Ste. 103, Vernon, CT 06066)

The Technical Committee on Architectural Acoustics (TCAA) is a Registered Provider in the American Institute of Architects (AIA) Continuing Education System (CES). The TCAA has developed a standardized introductory short course for architects. The TCAA short course is called "Architectural Acoustics" and attendance at this 1-hour-long course can earn an architect one continuing education unit (CEU). This paper will cover the administrative requirements of the AIA/CES, to prepare potential presenters. These requirements include the proper handling of paperwork, so that AIA members may receive credit for the course. Also, the manner in which the course is given is dictated by AIA requirements. TCAA membership and attendance at this workshop are required to qualify as an authorized presenter for this AIA/CES short course.

Session 3pAB

Animal Bioacoustics and Acoustical Oceanography: Autonomous Remote Monitoring Systems for Marine Animals IV

Catherine L. Berchok, Chair
National Marine Mammal Lab., Alaska Fisheries Science Ctr./NOAA, Seattle, WA 98115-6349

Invited Paper

1:00

3pAB1. Call localization of marine mammals using directional autonomous recorders. Susanna B. Blackwell, Charles R. Greene, Jr. (Greeneridge Sci., Inc., 1411 Firestone Rd., Goleta, CA 93117, susanna@greeneridge.com), Christopher S. Nations, Trent L. McDonald (WEST, Inc., Cheyenne, WY 82001), Aaron Thode (Univ. of California, San Diego, La Jolla, CA 92037), and A. Michael Macrander (Shell Exploration and Production Co., Anchorage, AK 99503)

Directional sensors of low-frequency acoustic waves have been used by navies in sonobuoys for submarine detection and localization for decades. Composed of an omnidirectional pressure sensor and two horizontal directional elements sensitive to particle motion, they provide information for determining the relative bearing to a sound source without ambiguity. They were adapted for bowhead whale monitoring in the mid-1980s and have been used in autonomous seafloor acoustic recorders since 2000. Applied to the coastal Beaufort Sea north of Alaska, fall bowhead migration was observed in detail during 2007 and 2008, providing a wealth of information on variability in the migration paths of calling whales and the influence of industrial sounds on the locations of calling whales. For example, many calls were detected within about 30 km of seismic survey activities, where received sound pressure levels from airgun pulses were often greater than 140 dB per 1 μPa. Also, seismic activities were correlated with statistically significant shifts in the whales' distance from shore, either offshore or inshore. However, interpretation of the results was challenged by the difficulty in distinguishing between a whale that ceases calling and a whale that deflects away from the study area. [Work supported by Shell Oil Co.]

Contributed Papers

1:20

3pAB2. Killer whale habitat use and prey fields from remote hydrophones and echosounders. Scott Veirs (Beam Reach Marine Sci. and Sustainability School, 7044 17th Ave NE, Seattle, WA 98115, scott@beamreach.org), Val Veirs (Colorado College, Colorado Springs, CO 80903), Jason Wood (The Whale Museum, Friday Harbor, WA 98250), Brian Moore, Bob McClure (BioSonics, Inc., Seattle, WA 98107), and Bob Otis (Ripon College, Ripon, WI 54971)

The Salish Sea Hydrophone Network has been providing live audio streams to listeners around the globe for the past three years. Five hydrophones around the core summer habitat of the endangered southern resident killer whales allow detection of killer whale presence and inference of pod identity based on call associations. Detection of calls, whistles, and clicks is accomplished automatically by spectral and time domain analysis and manually by a global network of trained human listeners. The hydrophone network is more effective than a sighting network for detecting killer whales during inclement weather and at night, and for generating real-time notifications to alert researchers and managers. Additionally, the live streams and archived recordings have enriched educational activities of local museums and aquariums. In 2008, the hydrophone at Lime Kiln State Park was supplemented by hydroacoustic surveys conducted using a Biosonics echosounder (200 kHz). From August through December the echosounder was deployed at 15 m depth to conduct a continuous, fixed, horizontal survey. Remote display technology was used to monitor the echosounder in real time while also listening to the live audio stream. The resultant time series of fish density and acoustic behavior of killer whales is presented.

1:35

3pAB3. Detecting humpback whale sounds in the Bering Sea: Confounding sounds in a cacophony of noise. Sara L. Heimlich, David K. Mellinger, Sharon L. Nieukirk, Holger Klinck (Cooperative Inst. for Marine Resour. Stud., Hatfield Marine Sci. Ctr., Oregon State Univ., 2030 S. Marine Sci. Dr., Newport, OR 97365), Kate Stafford, Sue E. Moore, and Phyllis J. Stabeno (Univ. of Washington, Seattle, WA 98105)

Humpback whales (*Megaptera novaeangliae*) are top predators of large zooplankton and forage fish, and one of the most common large whales in the Bering Sea. While present on feeding grounds, humpback whales produce nonsong sounds probably associated with feeding or social contacts. However, little is known about these highly variable sounds, and their detection is challenging. Recordings were collected during 2006–2007 at the long-term oceanographic moorings M2, M4, and M5 in the eastern Bering Sea. Passive acoustic detection of humpback whale calls in these recordings was confounded by a variety of other sounds, which fall within the same parameters as nonsong humpback vocalizations. An automatic algorithm that detects tonal sounds in the 300–950 Hz frequency band was used to find humpback calls. Raw detections were visually examined to verify the accuracy of the detections. This algorithm resulted in a significant number of wrong detections (false positives), especially sounds of bearded seals (*Erignathus barbatus*) and bowhead whales (*Balaena mysticetus*). An error rate was calculated to correct the high number of false detections, and further refinements were made to the automatic detection algorithm. The problems and possible solutions to detecting humpback whale sounds in complex acoustic recordings will be presented.

1:50

3pAB4. Monitoring cetaceans on seamounts in the Azores using passive acoustic techniques. Irma Cascão (Inst. of Marine Res., Dept. of Oceanogr. and Fisheries, Univ. of the Azores, Horta, Portugal, irmacascao@uac.pt), Marc O. Lammers (Univ. of Hawaii, Kailua, HI 96734), Monica A. Silva, Pedro Afonso, Rui Prieto, and Ricardo S. Santos (Univ. of the Azores, Horta, Portugal)

Seamounts are hotspots of marine life in the Azores, acting as feeding stations for diverse top predators, including cetaceans. Long-term monitoring of remote marine ecosystems is logistically difficult and expensive, involving high costs in terms of ship time and human resources, as well as being reliant on sea conditions. Passive acoustic methods are a highly efficient monitoring technique to study the distribution or abundance of

vocalizing cetacean species. In 2007, we began to deploy Ecological Acoustic Recorders (EARs)—two shallow (35 m) and two deep (190 m) units—around seamounts to explore and monitor the occurrence of cetaceans, their temporal patterns, and changing levels of activity. A wide variety of time series acoustic signals were collected, revealing the presence of different species of cetaceans and specific associations between species, based on their vocalization characteristics. Detections of vessels are being used to determine levels of activity at one of the study sites, Formigas bank marine reserve, making the EAR a useful tool to monitor protected areas. The EARs are an effective ecological instrument for monitoring cetaceans year-round and will provide a better understanding of the influence of seamounts on cetaceans' behavior.

WEDNESDAY AFTERNOON, 20 MAY 2009 PAVILION WEST, 2:15 TO 3:15 P.M.

Session 3pAO

Acoustical Oceanography: Acoustical Oceanography Prize Lecture

Jeffrey A. Nystuen, Chair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105-6698

Chair's Introduction—2:15

Invited Paper

2:20

3pAO1. Seabed characterization and model based processing: Past, present, and future. Martin Siderius (Dept. of Elect. and Comput. Eng., Portland State Univ., 1900 SW 4th Ave., Portland, OR 97201, siderius@pdx.edu)

One of the most active areas of research in acoustical oceanography has been the study of seabed acoustics. This has been partly motivated by the need for accurate sonar performance prediction, which often depends on knowledge of the seabed properties. There has also been significant research on environmentally adaptive and model based signal processing such as matched field processing. One of the obstacles to model based processing has been the required knowledge about seabed properties, which is needed as input. This dependency, in part, led to the innovation of matched field inversion methods to obtain seabed properties. In recent years, new algorithms have been introduced that use ambient noise (e.g., from breaking waves) as a surrogate sound source to probe the seabed and estimate geoacoustic properties. These noise methods have received much attention since measurements and processing are simple and the results are surprisingly good. In fact, current developments in noise processing may be the key to enabling model based methods that, in the past, were not considered practical. In this presentation a review of seabed inversion methods will be given with discussion about current developments in noise processing and the potential for renewed interest in model based processing.

Session 3pBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Shock Wave Therapy II

Michael R. Bailey, Cochair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105

Thomas J. Matula, Cochair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105

Contributed Paper

12:45

3pBB1. Interaction of single biological cell with tandem microbubbles in microfluidics. Georgy Sankin, Fang Yuan, and Pei Zhong (Dept. of Mech. Eng. Mat. Sci., Duke Univ., 144 Hudson Hall, Durham, NC 27708, gns @duke.edu)

Coupled oscillation of two laser generated microbubbles (maximum radius = 28 μm) and associated shear stresses are investigated experimentally. Bubble-bubble interaction in a microchannel of 25 μm height is observed using high-speed video cameras and μPIV technique. Two liquid microjets moving in opposite directions can be generated when the second bubble is produced at the maximum size of the first one. The interaction of these tandem microbubbles with single cell leads to controllable poration of adjacent cell membrane and dye uptake. μPIV data are compared with cell viability at various bubble-cell distances and azimuthal orientations. This method provides a new approach for highly selective cell treatment in situ, applicable to targeted microinjection of macromolecules and gene vectors in microfluidics devices. [Work supported in part by NIH.]

Invited Papers

1:00

3pBB2. Treating heterotopic ossification with shockwaves. Michael W. Chang (12340 NE 8th St., Ste. 101, Bellevue, WA 98005), Thanaphum Osathanon, and Cecilia Giachelli (Univ. of Washington, Seattle, WA 98195)

Heterotopic ossification (HO) is formation of lamellar bone in soft tissue, commonly seen among patients with burns, orthopedic fractures/procedures, brain/spinal cord injuries, muscle contusions, and rare hereditary disorders. Patients with HO experience pain and joint limitations, which further exacerbate their disability. Reduction of HO preventatively and/or for treatment has been attempted using medications, radiation, and surgery with limited success. Recent encouraging clinical HO outcomes using shockwave therapy have been discussed but treatment mechanisms were unclear. Mechanisms such as spallation, cavitation, squeezing, superfocusing, fatigue, and layer separation have been proposed for lithotripsy. Mineralized calcium phosphate macroporous nanofibrous fibrin scaffolds were used as a HO model to study treatment mechanisms. The HO was treated with electrohydraulic shockwave devices. The scaffold calcium contents decreased from 45.4(3.6) (untreated) to 20.9(18.1) (OssaTron, 0.71 mJ/mm2) and 27.5(23.5) ugCa/mg dry weight (EvoTron, 0.46 mJ/mm2) with 200 shocks (1 Hz). Scanning EM of the shockwave-treated HO supports separation of the calcium phosphate from the fibrin matrix probably contributed by shear and cavitation. Biological responses to shockwaves, such as inflammation, neovascularization, angiogenesis may also play important roles in clinical HO treatment. A unique clinical HO case is discussed to illustrate risks, benefits, and challenges from shockwave therapy.

1:15

3pBB3. Histotrispy: Urologic applications. William W. Roberts, Timothy L. Hall, Christopher R. Hempel (Dept. of Urology, Univ. of Michigan, TC 3879, 1500 East Med. Ctr. Dr., Ann Arbor, MI 48109-5330, willrobe@umich.edu), Zhen Xu, J. Brian Fowlkes, and Charles A. Cain (Univ. of Michigan, Ann Arbor, MI. 48109)

Histotripsy is an extracorporeal ablative technology that utilizes microsecond pulses of intense ultrasound to produce nonthermal, mechanical fractionation of targeted tissue. We have previously demonstrated the feasibility of histotripsy ablation in *in vivo* renal and prostate models. We sought to assess the chronic tissue response, tolerability, and safety of prostate histotripsy in a chronic *in vivo* canine model. Five acute and thirteen chronic canine subjects were anesthetized and treated with histotripsy targeting the prostate. Pulses consisting of three cycle bursts of 750-kHz ultrasound at a repetition rate of 300 Hz were delivered from a highly focused 15-cm aperture array. Prostates were harvested at 0, 7, 28, or 56 days after treatment. Transrectal ultrasound imaging provided accurate targeting and real-time monitoring of histotripsy treatment. Consistent mechanical tissue fractionation and debulking of prostate tissue was seen acutely and at delayed time points without collateral injury. Canine subjects tolerated histotripsy with minimal hematuria or discomfort. Only mild transient lab abnormalities were noted. Histotripsy is a promising noninvasive therapy for prostate tissue fractionation and debulking that appears safe and well tolerated without systemic side effects in the canine model.

1:30

3pBB4. Improving efficiency of ultrasound histotripsy sources by transmitting at the second harmonic frequency. Timothy A. Bigelow (Dept. of Elec. and Comput. Eng., Dept. of Mech. Eng., Iowa State Univ., 2113 Coover Hall, Ames, IA 50011, bigelow@iastate.edu)

Therapeutic ultrasound has shown potential for improving many aspects of medical care in recent years. Therapeutic ultrasound often involves high pressures or intensities, but pressures and intensities in tissue are limited by nonlinear propagation. The impact of nonlinear propagation is most severe for therapies involving cavitation where the peak rarefactional pressure plays the dominate role. As a result, in applications such as histotripsy, the size and power of the ultrasound source must be significantly increased in order to obtain the pressures needed *in vivo*. In this study, a new ultrasound source design was developed that will help mitigate the effects of nonlinear propagation. Specifically, sound was transmitted at both the fundamental and the second harmonic with the amplitude and phase of the second harmonic designed to counter nonlinear propagation effects for varying source powers from 10 to 500 W. The simulations demonstrated a 30% increase in the peak rarefactional pressure relative to a source where only the fundamental was transmitted when the appropriate phase was used for the second harmonic.

1:45

3pBB5. Skull flexure from blast waves: A mechanism for brain injury with implications for helmet design. William C. Moss, Michael J. King (Lawrence Livermore Natl. Lab., 7000 East Ave., Livermore, CA 94551, wmoss@llnl.gov), and Eric G. Blackman (Univ. of Rochester, Rochester, NY 14627)

Traumatic brain injury [TBI] has become a signature injury of current military conflicts. The debilitating effects of TBI are long-lasting and costly. Although the mechanisms by which impacts cause TBI have been well researched, the mechanisms by which blasts cause TBI are not understood. Various possibilities have been investigated, but blast-induced deformation of the skull has been neglected. From numerical hydrodynamic simulations, we have discovered that nonlethal blasts can induce sufficient flexure of the skull to generate potentially damaging loads in the brain, even if no impact occurs. The possibility that this mechanism may contribute to TBI has implications for the diagnosis of soldiers and the design of protective equipment such as helmets. [This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344].

2:00

3pBB6. Cavitation bubble activity during treatment of cortical bones by shock waves. Dorothée Bossis (CNRS LIP UMR 7623, 15 rue de l'Ecole de Medecine, 75006 Paris, France), Frederic Padilla, and Robin O. Cleveland (Boston Univ., Boston, MA 02114.)

Shock wave (SW) induced cavitation might be a physical effect by which bioeffects on bones are mediated, especially for SW treatment of bone nonunion. To assess the presence of cavitation and to study bubbles activity at the surface of cortical bones, we treated rat femurs with SWs using a clinical SW therapy device (Ossatron). Sequences of 14 images were acquired using a high speed camera capable of recording up to 100 million frames per second (Imacon). The sequences showed that for all voltage settings of the Ossatron a cloud of cavitation bubbles was induced at the surface of the bone. The bubbles near the surface of the bone appeared to form and act as a single larger bubble, which collapsed violently against the surface of the bone. For a voltage setting of 28 kV the collapse time was measured to be 720 micro-seconds after a single SW and reduced to 620 micro-seconds after 10 SWs at 1 Hz. In the case of a bone with an artificially fractured surface cavitation events were found to occur preferentially within the fracture gap. Cavitation might therefore be an important mechanisms in the SW treatment of bone defects.

2:15

3pBB7. Shock waves micro-damage induction in cortical bones: Comparison between experimental and simulations results. Frederic Padilla (Dept. of Mech Eng., Boston Univ., 110 Cummington St., Boston, MA 02115; CNRS LIP UMR 7623, 75006 Paris, France) and Robin O. Cleveland (Boston Univ., Boston, MA 02115)

Shock waves (SW) are considered a promising method to treat bone non unions. One potential mechanisms of action is the initiation of local microfractures, which may in turn trigger the start of bone healing. In this study, a set of eight intact rat femurs have been subjected to SWs (peak positive pressure of 40 MPa and peak negative pressure of -8 MPa, PRF of 2 Hz). The number of SW was varied from 50 to 1500. Micro-CT images of the specimens were acquired before and after treatment (16 microns resolution). In parallel, numerical simulations were used to quantify the stresses induced by SWs in cortical bone tissue. We used a 3-D FDTD code to solve the linear lossless equations that describe wave propagation in solids and fluids. A 3-D model of a fractured rat femur was obtained from micro-CT data. Results demonstrate that damages were induced in the bone tissues from 150 SWs. Comparison of the location of the induced damages on micro-CT images with predictions of maximum stresses by numerical simulations show that damages principally appeared in regions of highest stresses, suggesting that SW-induced tremendous stresses in bone tissue might be an important mechanisms implied in SW bone therapy.

2:30

3pBB8. Empirical angle-dependent tortuosity functions and sound transmission through cancellous bone. Keith Attenborough (Dept. of Eng., The Univ. of Hull, Cottingham Rd., Hull HU6 7RX, UK, k.attenborough@hull.ac.uk) and Haydar Aygun (Univ. of Hull, Hull HU6 7RX, UK)

Recent literature concerning the angle dependence of sound transmission through cancellous bone has suggested that it might be due either to elastic anisotropy or to microgeometrical anisotropy in the pore structure; e.g., an angle-dependent tortuosity. The elastic anisotropy approach has been found able to explain the observed variation in fast wave speed with angle better than the angle-dependent tortuosity at the cost of underpredicting slow wave speeds. In reality, it is likely that both influences are present in cancellous bone. Nevertheless the angle-dependent tortuosity approach has been revisited on the basis of orthotropic data for sound transmission through air-filled stereo-lithographical bone replicas [Attenborough *et al.*, JASA 118, 2779 (2005)]. [Work supported by Leverhulme Grant F/00181/N.]

2:45

3pBB9. Numerical simulation of propagation of acoustic and elastic waves in human head with fast integral equation solver. Elizabeth Bleszynski, Marek Bleszynski, and Thomas Jaroszewicz (Monopole Res., 739 Calle sequoia, Thousand Oaks, CA 91360)

An approach to numerical simulations of acousto-elastic waves in the human head is described, and simulation examples are presented. The primary goal of our work is to construct an efficient and high fidelity numerical simulation tool for investigating such effects as, e.g., acoustic energy transfer to the inner ear via non-airborne pathways. The solver employs a modified form of the volumetric Lippmann–Schwinger integral equation, adapted to high material-contrast problems, and an FFT based stiffness matrix compression implemented for distributed-memory systems enabling large scale (several million unknowns) simulations for realistic geometries. Some of the new aspects of the technique and its applications include (a) extension of the technique for problems involving large density contrasts,in situations involving elastic media; (b)improved, more economical, and more accurate Galerkin discretization scheme involving linear tetrahedral elements; (c) verification of the solver by comparison of its predictions with an exact analytical solution for a sphere composed with layers supporting both compressional and shear waves; (d) applications in representative numerical simulations of mechanisms of energy transfer to the inner human ear and

comparative analysis of the pressure distributions in a human head with and without the presence of noise protective devices, carried out with realistic geometry models. [Work supported by AFOSR].

3:00

3pBB10. Guided waves analysis using multiemitter and multireceiver arrays. Jean-Gabriel Minonzio, Marilyne Talmant, and Pascal Laugier (Universite Pierre et Marie Curie—Paris 6, Laboratoire d'Imagerie Paramétrique, CNRS LIP UMR7623, 75006 Paris, France)

Structural and material properties of elastic waveguides can be characterized by fitting measured to theoretical guided wave phase velocities. Here emitters and receivers are placed in contact on the same side of the waveguide (i.e., axial transmission geometry). Multireceiver arrays allow the determination of phase velocities using two-dimensional spatial-temporal Fourier transform, which requires a large distance probed by the receivers. Practical constraints, as in clinical inspection of cortical bones, may reduce the inspected spatial length and therefore the efficiency of this technique. We propose a technique which takes benefit of using both multiple emitters and multiple receivers. This approach derived from the DORT method adapted to the transmission problem. The singular values decomposition at different frequencies of the transmission matrix between the arrays give the invariants of transmission, which are linked with the guided modes. First, experiments carried out on different metallic plates are shown. Experimental velocities are in good agreement with Lamb waves theoretical values. In particular, zero group velocity resonance and negative phase velocities are shown. These results allows a good evaluation of the thickness and the transverse and longitudinal bulk waves velocities. Further applications concerning evaluation of elastic properties of cortical bone are finally mentioned.

WEDNESDAY AFTERNOON, 20 MAY 2009 EXECUTIVE SALON II/III, 1:00 TO 2:00 P.M.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

Paul E. Barbone, Chair
Dept. of Aerospace and Mechanical Engineering, Boston Univ., Boston, MA 02215

Invited Papers

1:00

3pID1. Hot topics in youth education outreach. Grace Klonoski (Optical Society of America, 2010 Massachusetts Ave., N.W., Washington, DC 20036-1012)

This presentation will focus on ways in which professional societies and their volunteers can make a lasting, positive difference in supporting the science and math learning of young students. The introduction will provide a brief overview of trends in student achievement in math and science and will discuss the availability of qualified teachers in the United States. Four features of effective learning environments will be described. The remainder of the presentation will be devoted to examples of how scientific communities can become directly involved in reaching out to students on a local and global scale. A variety of youth programs hosted by the Optical Society of America (OSA) and the OSA Foundation will be highlighted with an emphasis on the scope, goals, and quantitative results of each initiative. From a multilingual Web site that provides students, teachers, and parents with basic science information, lesson plans, and hands-on activities, to the production and worldwide distribution of "optics suitcases" that give students the opportunity to participate in hands-on demonstrations, OSA's recent efforts have benefited thousands of individuals in more than 20 countries. The planned collaboration between ASA and OSA to develop acoustics-focused classroom materials, a Web site, and demonstration kit will be described.

1:20

3pID2. Auditory attention and the active listener. Barbara Shinn-Cunningham (Cognit. and Neural Systems, 677 Beacon St., Boston, MA 02215)

Traditionally, psychoacousticians and auditory physiologists have focused on how stimulus properties affect perceptual abilities, measuring just-detectable differences in different acoustic attributes or thresholds for audibility of different sounds. However, in everyday settings with more than one audible sound source, these low-level abilities are often not the factors limiting understanding of an important sound source. Instead, the way the brain organizes sound and focuses attention has an enormous impact on what we actually hear and analyze out of an auditory scene. Recent studies in both perceptual and physiological acoustics are working to uncover how the intentions and goals of the listener influence processing of sound in the brain. This talk will review some recent advances in auditory neuroscience that demonstrate the importance of top-down attention on neural processing and perception of sound.

1:40

3pID3. Hot topics of signal processing in acoustics. Ning Xiang (Graduate program in Archit. Acoust., Rensselaer Polytech. Inst., Troy, NY, 12180) and David Chambers (Lawrence Livermore Natl. Lab., Livermore, CA 94551)

Signal processing is used to some extent in all areas of acoustics, such as extracting relevant information from acoustic measurements made either in the laboratory or in the field, processing signals and/or synthesizing data to cope with demanding tasks raised in acoustics. Techniques range from simple classical approaches based on Fourier transforms and Gaussian noise, to sophisticated model-based techniques that incorporate physical/parametric models of the acoustical system. In this paper we highlight new approaches to

signal processing that could be applied to a broad variety of acoustical problems. These include coded signals for architectural-acoustics, acoustical communications, and medium characterization, Bayesian methods for room acoustics, physical acoustics, and underwater acoustics including highly nonlinear problems with non-Gaussian noise, and extensions to the familiar Kalman filtering to nonlinear models. Examples of each approach will be shown that illustrate the advantages and disadvantages of each technique. Additional topics may be discussed as time allows.

WEDNESDAY AFTERNOON, 20 MAY 2009 PARLOR B/C, 1:00 TO 2:30 P.M.

Session 3pMUa

Musical Acoustics: Acoustics of Bagpipes

D. Murray Campbell, Chair
School of Physics and Astronomy, Univ. of Edinburgh, Edinburgh EH9 3JZ, U.K.

Chair's Introduction—1:00

Invited Papers

1:05

3pMUa1. An introduction to bagpipes of the world. Paul A. Wheeler (Dept. of Elec. & Comput. Eng., Utah Stat Univ., 4120 Old Main Hill, Logan, UT 84322-4120)

When one hears the term "bagpipe," the Great Highland bagpipe from Scotland comes to mind. There are actually several types of bagpipes played across Europe and the Middle East, each with its unique characteristics. This paper classifies bagpipes from around the world based on their construction and acoustic properties. These include: method of filling the bag (bellows versus mouth pipe), bag properties and materials, reeds (single or double), pipes used for chanters and drones (cylindrical or conical), and general characteristics of the instrument. The objective of this overview paper is to provide the background for more detailed papers in the session on bagpipes.

1:30

3pMUa2. A brief history and acoustical analysis of the great highland bagpipe. Stanley A. Cheyne (Dept. of Phys. and Astronomy, Hampden-Sydney Coll., Hampden-Sydney, VA 23943)

A brief history and analysis of the great highland bagpipe will be discussed. Although the bagpipe may have originated several thousands of years ago, the great highland bagpipe, distinguished by three drones and a chanter, is only a few hundred years old. A discussion of the components and operation will be given. In addition, spectral analysis of the drones and chanter will be discussed and how the harmonic content of the drones support the notes played on the chanter. A spectral analysis and sound pressure levels of chanter reeds of varying ages, strengths, and materials will also be presented.

2:00

3pMUa3. Tuning and tone quality of bagpipe drones. R. Dean Ayers (Phys. Dept., Southern Oregon Univ., 1250 Siskiyou Blvd., Ashland, OR 97520, ayersr@sou.edu) and Peter R. Nordquist (Southern Oregon Univ., Ashland, OR 97520)

Tunes are played on the Great Highland bagpipe using the tone holes of its conical chanter. The much quieter practice chanter has holes with the same spacings on a narrow cylindrical bore. Differences between spherical waves in a cone and plane waves in a cylinder give rise to striking differences in pitch and tone quality. Drone pipes do not need to fit into either the conical family or the closed-pipe family of reed woodwinds, because each drone produces only one note. A bass drone has three cylindrical segments, with diameters that increase in sequence, followed by an expansion chamber (bell and constricting cap) at the output end. Tuning slides between the cylinders create two additional expansion chambers with large diameters and adjustable lengths. (A tenor drone has two cylinders and one tuning slide.) Sound waves incident from either direction experience strong reflections at each jump in diameter, resulting in uneven standing waves and irregular input impedance curves that would be impractical for pipes with tone holes. Experimental results and computer models are used to relate the shape of a drone's air column to its tuning and tone quality. [Work supported by the Paul S. Veneklasen Research Foundation.]

Session 3pMUb

Musical Acoustics: Bagpipe Lecture and Concert

D. Murray Campbell, Chair
School of Physics, Univ. of Edinburgh, Edinburgh EH7 3JE, UK

Kevin Carr, a bagpipe performer will demonstrate and discuss various bagpipes from around the world.

Session 3pPA

Physical Acoustics: Influence of Temperature on Sound in Condensed Matter

Albert Migliori, Chair
Los Alamos National Lab., Los Alamos, NM 87545

Chair's Introduction—12:55

Invited Papers

1:00

3pPA1. Precise sound velocity measurements on solids and liquids at high pressure and high temperature with direct length measurement. Baosheng Li (Mineral Phys. Inst., Stony Brook Univ., Stony Brook, NY 11794, Baosheng.Li@sunysb.edu)

Sound velocity measurements at pressure and/or temperature provide important information in the material's structural behavior and physical properties under extreme conditions. With the application of synchrotron X-radiation sources, new developments have emerged to facilitate the study properties and structures of crystalline and amorphous materials. In the last decade, we have developed a unique technique for measuring sound velocity in a cubic type multianvil apparatus installed at X-17B2, National Synchrotron Light Source (NSLS) of Brookhaven National Lab, which allows us to conduct precise velocity measurements on crystalline and noncrystalline materials at various pressure and temperature conditions by conducting simultaneous X-radiation and ultrasonic interferometry measurements. In this paper, sound velocities at high temperatures for some ceramics and metals are measured to study the behavior and property change of these materials undergoing compression/expansion as well as phase transition and structural instabilities. This technique opens new opportunities for the study of materials that are of interest to many disciplines, such as earth science, materials science, and condensed matter physics.

1:25

3pPA2. The influence of grain structures on resonance behavior near phase transitions. Timothy W. Darling, Gunes Kaplan (Phys. Dept.-NTF Univ. of Nevada, 1664 N. Virginia St., Reno, NV 89557), James A. TenCate (Los Alamos Natl. Lab., Los Alamos, NM 87545), Ruth E. A. McKnight, and Michael A. Carpenter (Univ. of Cambridge, Downing St., Cambridge, UK)

Inhomogeneities such as grains or plastically deformed regions in materials locally change sound wave velocities and introduce scatter in the pattern of expected resonance frequencies. The accompanying variation in the components of the free energy may promote, or block, nucleation and growth of new phases near transition temperatures. Lanthanides and martensites are two groups of materials with complex phase structures where control of the phase transformation is important in technological applications. Complex oxide minerals with a wide variety of phase structures are important in understanding the behavior of the earth's crust. The influence on mechanical resonances of both the micro (meso) structures and the appearance and growth of phases near transition temperatures will be discussed. Experimental data on resonance behavior in inhomogeneous materials (largely lanthanides, transition metals, and silicates), where attempts have been made to control the scale and type of microstructure as they approach phase transitions will be presented. [This work receives support from DOE Grant No. DE-FC52-06NA27616 through the University of Nevada Terawatt Facility.]

1:50

3pPA3. Temperature dependence of elastic moduli for polycrystalline Pu: Comparison of alpha and gamma phases. I.R. Stroe (Worcester Polytechnic Inst., 100 Institute Rd., Worcester, MA 01609, izabela@wpi.edu), J. B. Betts, C. Pantea, A. Trugman, J. N. Mitchell, M. Ramos, F. Freibert, C.H. Mielke, and A. Migliori (Los Alamos Natl. Lab., Los Alamos, NM 87544)

The elastic properties of pure polycrystalline Pu were investigated as a function of temperature by resonant ultrasound spectroscopy. For both alpha and gamma phase, shear (G) and longitudinal (CL) elastic moduli were determined simultaneously and the bulk modulus (B) was computed from them. In alpha phase, G and B exhibit the same temperature dependence, but an unusually large change. This behavior together with a high Gruneiser parameter and an almost temperature independent Poisson ratio is attributed to the 5-f electron localization delocalization. In the gamma phase, G and B show anomalous temperature behavior: G decreases with temperature while B is near temperature independent. The Poisson ratio increases with temperature by 4%. This normal behavior of the Poisson ratio is unusual for Pu.

Contributed Papers

2:15

3pPA4. Stability of monitoring weak changes in multiply scattering media with ambient noise correlation: Laboratory experiments. Larose Eric, Hadziioannou Celine, Coutant Olivier, Roux Philippe, and Campillo Michel (Lab. de Geophysique interne et Tectonophysique, CNRS & Univ. J. Fourier, Grenoble, France)

Previous studies have shown that small changes can be monitored in a scattering medium by observing phase shifts in the coda. Passive monitoring of weak changes through ambient noise correlation has already been applied to seismology, acoustics, and engineering. Usually, this is done under the assumption that a properly reconstructed Green function as well as stable background noise sources are necessary. In order to further develop this monitoring technique, we performed a laboratory experiment in the 2.5 MHz range in a gel with scattering inclusions, comparing an active (pulse-echo) form of monitoring to a passive (correlation) one. Our results show that temperature changes in the medium can be observed even if the Green function (GF) of the medium is not reconstructed. Moreover, we establish that the GF reconstruction in the correlations is not a necessary condition: the only condition to monitoring with correlation (passive experiment) is the relative stability of the background noise structure.

2:30

3pPA5. Temperature variation of Young's modulus and elastic wave speed in a mixture of solid-particles in liquid. Hasson M. Tavossi (Dept. of Phys., Astronomy and Geosciences, Valdosta State Univ., 1500 No. Patterson St., Valdosta, GA 31698)

The variation of ultrasonic wave speed in the solid-liquid mixtures as a function of temperature is studied. The goal of this investigation is to determine temperature dependence of the elastic moduli of materials consisting of different size solid particles. The elastic wave speed in solids is assumed to be proportional to the square root of Young's modulus of the material. Elastic wave speed in the material is related to the variation of Young's modulus of the solid constituents. Also considered are temperature dependence of the wave attenuation, by scattering and absorption, and the effects of phonon modes, their amplitude and frequency, on the elastic moduli and wave dispersion in the material. Previous experimental data have shown a decrease in elastic moduli at the contact points of solid particles with increase in elastic wave frequency. Findings on the variations of elastic wave speed in solids with temperature and frequency are compared to analyze any similarity between the effect of wave frequency increase and temperature increase.

2:45

3pPA6. Resonant ultrasound spectroscopy studies of Fe-based superconductors. Veerle Keppens, Lindsey VanBebber, Yanbing Luan (Dept. Mater. Sci. and Eng., The Univ. of Tennessee, Knoxville, TN 37996), Michael A. McGuire, Athena Sefat, Brian C. Sales, and David Mandrus (Oak Ridge Natl. Lab., Oak Ridge, TN 37831)

The discovery of high superconducting transition temperatures in Fe-based compounds has generated a frenzy of experimental and theoretical research activities on these systems. Whereas doping is required to establish superconductivity, careful studies of the undoped materials are important to elucidate the underlying physics. Recently, resonant ultrasound spectroscopy measurements were carried out on the undoped oxypnictide system LaFeAsO as well as the oxygen-free systems $BaFe_2As_2$ and $Ba(Fe, Co)_2As_2$. The elastic response of the parent compounds shows a very gradual softening corresponding to the structural transition that characterizes these systems. Remarkably, this softening persists in the superconducting samples, even though the structural transition in these compounds is suppressed.

Session 3pSC

Speech Communication: Lexical Effects and Perceptual Processing in Speech (Poster Session)

Eric A. Vatikiotis-Bateson, Chair
Dept. of Linguistics, Univ. of British Columbia, Vancouver, BC V6T 1Z1, Canada

All posters will be on display from 1:00 p.m. to 3:20 p.m. To allow contributors an opportunity to see other posters, all contributors of odd-numbered papers will be at their posters from 1:00 p.m. to 2:10 p.m. and contributors of even-numbered papers will be at their posters from 2:10 p.m. to 3:20 p.m.

Contributed Papers

3pSC1. Phonetic experience with specific words affects categorical perception of those words. Mark VanDam (Boys Town Natl Research Hospital, 555 N. 30 St., Omaha, NE 68131, vandamm@boystown.org) and Robert F. Port (Indiana Univ., Bloomington, IN 47405)

How does experience with specific words influence linguistic-phonetic categories? Listeners were trained over a five-session, listen-and-repeat task on a set of target words embedded in continuous speech and altered so that the initial stop consonant voice-onset time (VOT) was 80% longer than natural. Voicing boundaries were estimated before and after training using a two-alternative, forced-choice perceptual task on an eight-step VOT continuum. Stimuli were highly natural tokens by two stimulus talkers. Part 1 of the experiment asked whether exposure to lengthened forms would influence location of the voicing boundary, and, if so, whether that effect would generalize to similar forms. Results showed longer boundaries after exposure to lengthened VOTs for the trained forms, but lengthening did not generalize to new forms. Part 2 investigated voicing boundary locations as a function of lexical status (word, nonword) and usage frequency (high, low). Boundary locations indicated expanded VOT regions both for nonwords over words (opposite to the Ganong effect) and for high-frequency words over low-frequency words; neither lexical status nor usage frequency interacted with training. Results suggest a lexical sensitivity to low-level speech cues, thus offering support for a rich memory language model.

3pSC2. Were we or are we? Perception of reduced function words in spontaneous conversations. Natasha Warner, Dan Brenner, Anna Woods (Box 210028, Univ. of Arizona Dept. of Linguist., Tucson, AZ 85721-0028), Bejamin V. Tucker (Univ. of Alberta, Edmonton, AB, Canada T6G 2E7), and Mirjam Ernestus (Univ. Nijmegen, Nijmegen AH6500, the Netherlands)

Spontaneous, reduced pronunciations diverge greatly from citation forms. The quality of a single segment can change, e.g., /b/ in "about" surfacing as an approximant. But sounds, syllables, and entire words can also be deleted (e.g., "do you have time?" as [djutEm] with no acoustic trace of "have"). This work investigates the perception of reduced function words such as "he was" or "we were." Twenty-two young American English speakers' spontaneous conversations with close acquaintances were recorded. From these, we selected utterances containing items such as "he's/he was, we're/we were, got him/got them." When hearing an entire utterance, native listeners may clearly perceive "we were," but on hearing just the "we were" portion, they perceive an unambiguous "we're." The portion of the signal presented to listeners is manipulated to determine the contributions of local acoustic cues, speech rate and coarticulation, semantic and syntactic information, and overall bias toward present vs past tense. An auditory and a written task are also compared to separate the contribution of intonation from that of syntax/semantics. These results begin to elucidate the interplay of information sources listeners draw upon when parsing spontaneous speech. Future work will compare to non-native listeners' perceptions.

3pSC3. The roles of tone and syllable structure in Mandarin spoken word recognition. Yuwen Lai (Dept. of Linguist., Univ. of British Columbia, 2613 West Mall, Vancouver, BC V6T 1Z4, Canada) and Jie Zhang (Univ. of Kansas, Lawrence, KS 66044-3177)

The present study adopts the gating paradigm to investigate the roles of tone, onset sonorancy, and nasal coda in Mandarin spoken word recognition. Duration-blocked gates generated from eight monosyllabic quadruplets with matching frequencies of occurrence were used as stimuli. The initial consonant of each syllable formed the first gate, with later gates formed by 40 ms increments. Twenty-eight native Mandarin speakers from Beijing were asked to identify each gated stimulus by writing down the Chinese characters. Isolation point (IP) based on correct tone identification as well as overall correct word identification (correct onset, rhyme, and tone) were collected. Results from both conditions showed that tone 1 has an earlier IP than tone 4, which has an earlier IP than tones 2 and 3. Sonorant-initial syllables have an earlier IP than obstruent-initial syllables, but further analyses of covariance indicated that this is due to the fact that IP covariates with the duration of the initial consonant. Syllables without a nasal coda have an earlier IP than syllables with a nasal coda. This effect might be due to the interference of nasalization on tone perception or the delayed tonal contour realization due to the nasal coda [Xu, (1998)].

3pSC4. Duration and context speech rate as cues to lexical perception and word segmentation. Molly J. Henry, Laura C. Dilley, Louis N. Vinke, and Christopher J. Weinland (Dept. of Psychol., Bowling Green State Univ., Bowling Green, OH 43403, mjhenry@bgsu.edu)

Duration and speech rate are traditionally assumed to be filtered out before lexical lookup takes place, although these factors are known to influence phoneme perception. Here, the hypothesis was investigated that duration can affect both perceived lexical identity, as well as the perceived number and implied locations of word boundaries relative to the speech signal. Experiment 1 was a production study which investigated durations of vocalic portions of phonetically similar versions of target word strings which differed in their number of syllables (e.g., cease versus see us); these target word strings were spoken in semantically neutral context sentences. As expected, vocalic durations in target strings with fewer syllables were shorter than those with more syllables. In Experiment 2, the relative durations of vocalic portions of target strings in sentences from Experiment 1, as well as sentential context speech rate, were manipulated using speech resynthesis. Relative duration and context speech rate both affected the words that participants heard, as well as the implied number of phonemes and imputed locations of word boundaries. These findings indicate that duration plays a significant and underinvestigated role in spoken word recognition and word segmentation.

3pSC5. The spread and density of the phonological neighborhood can strongly influence the verbal transformation illusion. James A. Bashford, Jr., Richard M. Warren, and Peter W. Lenz (Dept. of Psychol., Univ. of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201)

When a recorded verbal stimulus repeats over and over, adaptation occurs and listeners hear competing forms. Reports of these "verbal transformations" (VTs) were obtained for 36 consonant-vowel (CV) syllables that varied both in frequency-weighted neighborhood density (ranging from 12.73 to 90.42) and in neighborhood spread [i.e., for 18 CVs, changes at either phoneme position could yield real words (spread = 2) while for the remaining 18 CVs, changes at only one position could yield words (spread

$= 1)$]. The strength of the VT illusion, measured by the amount of time the stimuli were heard nonveridically during the 300-s repetition period, decreased substantially with both increasing neighborhood density [$r=-0.74$, $F(1,34)=42.6$, $p<0.0001$] and increasing spread [$r=-0.75$, $F(1,34) = 44.1$, $p<0.0001$]. Stepwise regression revealed that density and spread collectively accounted for approximately 70% of the variance in illusion strength [$F(1,33)\geqslant 10.0$, $p<0.003$ or better]. These effects are larger than, but generally consistent with, neighborhood effects obtained with other psycholinguistic tasks, and they suggest that VTs can provide a highly sensitive measure of lexical competition. [Work supported by NIH.]

3pSC6. Using auditory feedback during vocalization: Different mechanisms for utterance onset and pitch maintenance. Colin S. Hawco and Jeffery A. Jones (Dept. of Psych., Wilfrid Laurier Univ., 75 University Ave. West, Waterloo, ON N2L 3C5, Canada)

During vocalization, auditory feedback is used to attain and maintain a desired fundamental frequency (F0). The use of auditory feedback to control vocalization has been studied using adaptation studies, in which an F0 feedback alteration is maintained across many trials, and using midutterance perturbations, where a brief F0 change is introduced during an ongoing utterance. Participants vocalized during two blocks. In an adaptation block, a one semitone F0 shift was maintained for 40 trials, and midutterance perturbations were introduced by removing the feedback alteration. In a control block, a midutterance perturbation was introduced in trials, which were either unshifted or shifted randomly by one semitone at before the onset of vocalization (similar to the adaptation trials). F0 compensation to midutterance perturbations was identical in all conditions, and was always smaller than the compensation to a shift at utterance onset. These results are explained by a change in the control strategy at utterance onset and midutterance. At utterance onset, auditory feedback is compared to feedback predicted by a forward model to ensure the pitch goal is achieved. However, after utterance onset, the control strategy switches and stabilization is maintained by comparing feedback to previous F0 production.

3pSC7. Auditory-phonetic projection and lexical structure in the recognition of sine-wave words. Robert E. Remez, Kathryn R. Dubowski, Robin S. Broder, Morgana Davids, Yael S. Grossman, Marina Moskalenko, Jennifer S. Pardo, and Sara Maria Hasbun (Dept. of Psych., Barnard College, Columbia Univ., 3009 Broadway, New York, NY 10027, remez @columbia.edu)

Speech remains intelligible despite the elimination of canonical acoustic correlates of phonemes from the spectrum. Listeners tolerate distortion or spectral blur in tone analogs, noise band vocoded speech, and acoustic chimeras in utterances ranging from syllables to isolated words and sentences. A portion of this flexibility is attributable to short-term perceptual learning in auditory-to-phonetic projection, though exploiting the properties of lexical neighborhoods plays a role with utterances composed of words. New tests were conducted to estimate talker learning, segmental sensitivity, and lexical knowledge in this kind of perceptual versatility. Sine-wave versions of the easy/hard word lists were created, and the performance-level difference between the two lists was used to index the default reliance on lexical processes. Several kinds of preliminary exposure were used to induce sensitivity to a dimension of perceptual learning: sine-wave speech produced by the same talker, sine-wave speech of a different talker, natural speech of the same talker. A comparison of exposure effects on the performance level of easy/hard word recognition offered a clue about the differential contribution of talker-based, segment-based, and lexically-based attention in speech perception without canonical spectra. Implications for perceptual accounts based on cue likelihood will be discussed. [Work supported by NIDCD.]

3pSC8. Function words of lexical bundles: The relation of frequency and reduction. Shannon F. Lemke, Antoine Tremblay, and Benjamin V. Tucker (Dept. of Linguist., Univ. of Alberta, 4-32 Assiniboia Hall, Edmonton, AB, Canada, T6G 2E7, sflemke@ualberta.ca)

Studies of spontaneous speech have shown frequency effects on the amount of reduction produced by speakers, demonstrating that predictability facilitates production of a target word [Shi *et al.* (2005); Jurafsky *et al.* (2001); Bell *et al.* (2003)]. This paper investigates the amount of reduction produced in laboratory recorded speech and considers the effect of frequency on the duration of function words in four-word sequences. It is also found that the influence of frequency has an effect on holistically storing these bundles. An interaction between word position and the third-order transitional probability (ABC $\rightarrow$ D) has been established, indicating that greater third-order transitional probabilities predict shorter function word durations in the first and second positions of a bundle, and, therefore, involve more durational reduction. The current research shows that, just as frequency affects reduction in spontaneous speech, there is an effect in laboratory produced speech as well. These findings indicate that multiword sequences are stored as lexical units.

3pSC9. Basic word segmentation abilities emerge earlier in infancy than previously thought. Elizabeth K. Johnson (Dept. of Psych., Univ. of Toronto Mississauga, 3359 Mississauga Rd. North, Mississauga, ON L5L 1C6, Canada, elizabeth.johnson@utoronto.ca) and Amanda Seidl (Purdue Univ., West Lafayette, IN 47907)

English-learning 7.5- but not 6-month-olds extract word forms from fluent speech (Jusczyk *et al.*, 1999). Thus, English learners are thought to begin segmenting words from speech by 7.5 months. However, recent research has shown that when target words are flanked by a frequent and emotionally salient word (e.g., the infant's name), even 6-month-olds can extract words from speech (e.g., Bortfeld *et al.*, 2005). This suggests that basic word segmentation capabilities may emerge earlier than past studies have suggested. Using the Headturn Preference Procedure, we tested 6-month-olds' ability to segment utterance-flanked words from speech, e.g., "geff" from "At the circus we met a silly geff." Infants were familiarized with passages containing target words in utterance initial and final position, and then tested on their recognition of these words in isolation. A significant looking time difference to familiar versus unfamiliar words was found, indicating that 6-month-olds segmented the target words from speech. Six-month-olds' success at segmenting utterance-flanked words from speech is particularly interesting because infant-directed speech consists of short utterances containing many utterance-flanked words. Segmentation of utterance-flanked words could help infants learn the cues needed to extract harder utterance-medial words from speech (see Seidl and Johnson, 2006).

3pSC10. Phonological competition during the recognition of spontaneous speech: Effects of linguistic context and spectral cues. Susanne Brouwer, Holger Mitterer, and Falk Huettig (MPI for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen)

How do listeners recognize reduced forms that occur in spontaneous speech, such as "puter" for "computer"? To this end, eye-tracking experiments were performed in which participants heard a sentence and saw four printed words on a computer screen. The auditory stimuli contained canonical and reduced forms from a spontaneous speech corpus in different amounts of linguistic context. The four printed words were a "canonical form" competitor (e.g., "companion", phonologically similar to "computer"), a "reduced form" competitor (e.g., "pupil", phonologically similar to "puter"), and two unrelated distractors. The results showed, first, that reduction inhibits word recognition overall. Second, listeners look more often to the "reduced form" competitor than to the "canonical form" competitor when reduced forms are presented in isolation or in a phonetic context. In full context, however, both competitors attracted looks: early rise of the "reduced form" competitor and late rise of the "canonical form" competitor. This "late rise" of the "canonical form" competitor was not observed when we replaced the original /p/ from "puter" with a real onset /p/. This indicates that phonetic detail and semantic/syntactic context are necessary for the recognition of reduced forms.

3pSC11. Lexical access across different voices. Rochelle S. Newman (Dept. Hearing & Speech Sci., and Program in Neurosci. and Cognit. Sci. Univ. of Maryland, 0100 Lefrak Hall, College Park, MD 20742, rnewman @hesp.umd.edu)

In many real-world situations, listeners may be trying to attend to one individual while other voices are speaking in the background. These separate streams of sound hit the ear simultaneously. Within a single stream, research has investigated the extent to which lexical activation is limited by word boundary information. For example, Gow and Gordon (1995) found that multiple-word sequences such as "two lips" may activate words such as "tulips" despite the presence of acoustic cues to a word boundary in the middle of the sequence. Yet, no studies have examined whether lexical ac-

cess could "cross" signals entirely. In a series of four experiments, we test whether lexical access might (ever) cross talker boundaries. We find that when listeners hear two words spoken by two different voices (for example, a male saying "two" followed by a female saying "lips") they still activate the larger cross-voice word "tulips," despite the presence of strong cues to an acoustic boundary. [Work supported by NSF.]

3pSC12. Speaking rate modulates lexical competition in online speech perception. Eva Reinisch, Alexandra Jesse (Max Planck Inst. for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The Netherlands, Eva.Reinisch@mpi.nl), and James M. McQueen (Radboud Univ. Nijmegen, and Max Planck Inst. for Psycholinguistics, Nijmegen, The Netherlands)

Durational cues are used in online word recognition to disambiguate Dutch phrases such as "een(s) (s)peer" ["on(c)e (s)pear"]: The longer the [s], the more likely listeners consider the target to be [s]-initial [K. E. Shatzman and J. M. McQueen, Perc. Psychophys. **68**, 1–16 (2006)]. Here, a series of eye-tracking experiments show that speaking rate affects perception of duration and thus modulates lexical competition online. Listeners' eye-movements were recorded while they listened to sentences such as "Ze heeft wel eens peer gezegd" ("She once said pear"). Listeners had to click on printed targets (peer) presented on a computer screen with an [s]-initial competitor (speen; "pacifier") and two unrelated distractors. When the preceding context is faster, the critical [s] should sound longer and match better the beginning of "speen." Listeners indeed looked at the [s]-initial competitors more following fast than slow contexts. The opposite was found for [s]-initial targets (sneeuw; "snow") with non-[s]-initial competitors (neef; "nephew"). Context adjacent to the target (i.e., wel eens) influenced target perception more than nonadjacent context. Nonadjacent context, however, was sufficient to modulate competition when adjacent context was spoken at the same rate as the target. Amount of nonadjacent context at a given rate affected only off line categorization.

3pSC13. The nonaccommodation of speech errors. Andrea Gormley (Inst. of Cognit. Sci. Rm. 2222, Dunton Tower, Carleton Univ., 1125 Colonel By Dr. Ottawa, ON K1S 5B6, Canada, agormley@connect.carleton.ca)

Word-form speech errors are assumed to adapt to the unintended environment. For example, in the error 'bads cat' for intended 'bad cats,' the plural marker 's' assimilates to the voicing of the 'd' in 'bad.' This phenomenon, called accommodation, provides evidence that the component responsible for errors is processed before the phonological assimilation component. Previous work on accommodation relies on the researcher to detect accommodation [D. Boomer and J. Laver, Br. J. Disord. Commun. **3**, 1–12 (1968)]. Given that these studies are prone to perceptual bias, the conclusion that accommodation is the norm remains open. An acoustic analysis of errors was conducted to re-address this question. Thirty-two nonword tongue twisters, e.g., 'tiff tivv tivv tiff,' were designed to induce voicing errors on the coda. Because vowel lengthening before voiced codas is a phonological process in English, vowel length can be measured to detect accommodation. Errors were determined for each participant by measuring coda percent voicing and vowel duration in a control condition. Results from six participants (872 errors) show that while errors do accommodate (7.6%), nonaccommodation (40.7%) occurs more frequently. This result shows that errors do occur after phonological processing. [Work supported by OGS and Carleton University.]

3pSC14. Planning time effects of phonological competition: Articulatory and acoustic data. Christine R. Mooshammer (Haskins Labs., 300 George St., New Haven, CT 06511, tine@haskins.yale.edu), Louis Goldstein (USC, CA), Mark Tiede (MIT Res. Lab of Electron., MA), Manisha Kulshreshtha (Haskins Labs.,), Scott McClure, and Argyro Katsika (Haskins Labs.,)

One major cause for speech production variability and errors is competition between phonologically similar sequences in an utterance. Since one recent model of speech production planning [Nam (2004)] also posits a systematic relation between planning time and kinematic variability, we decided to directly investigate whether competition increases planning time, i.e., whether it takes longer to initiate a sequence such as "tape cape" compared with "tape tape." Effects of competition in the onset were compared to competition in the coda (e.g., "tape take"). Results from two studies are reported: articulatory latencies from a delayed naming task recorded using EMA (four speakers), and acoustic latencies from a delayed naming task, a simple naming task, and a picture naming task (ten speakers). Latencies were significantly affected by competition, i.e., latencies were longer for items like "tape cape" and "tape take" than for "tape tape." However, no significant differences were found in competition effects between onset and coda positions. Apart from latency, overall duration increased and clusters formed across words showed less overlap. Observed effects on latency were largest for picture naming, followed by simple naming and least for delayed naming. [Work supported by NIH DC008780.]

3pSC15. Color effects in audiovisual spoken word recognition. Rachael Frush Holt, Tessa Bent (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S Jordan Ave., Bloomington, IN 47405, raholt@indiana.edu), and Luis Hernandez (Indiana Univ., Bloomington, IN 47405)

Although the importance of the visual contribution to speech perception has been known for over 50 years, strong interest in audiovisual speech perception has emerged only recently. In developing a test of audiovisual sentence recognition, white balancing and color correcting (adjusting the whites, mids, and black bandwidths to compensate for the camera's inability to fully reproduce human eyesight color spectrum) improved visual-only spoken word recognition. No differences were found in audiovisual or auditory-only modalities. Although improvement was seen across all key words, those with more visible phonemes showed the greatest improvement. To further examine the effects of color in the audiovisual modality specifically, the original speech signal was replaced with a signal-correlated noise and presented to adults in one of three conditions: non-color-corrected, color-corrected, and color-inverted. Minimal speech cues were sufficient for overcoming performance decrements due to small, but not large, color disturbances. Furthermore, in the color-corrected condition, performance in visual-only and signal-correlated noise conditions were equivalent, whereas in the non-color-corrected condition, signal-correlated noise improved performance over visual-only. The results suggest that consideration of minor color disturbances is warranted if testing visual-only perception; however, significant color disturbances can cause decrements in audiovisual conditions, as well. [Work funded by the AHRF.]

3pSC16. Perception of voiced-only and noise-vocoded speech by a language-trained chimpanzee (*Pan troglodytes*). Lisa A. Heimbauer (1527 Pangborn Station Dr. Decatur, GA 30033, lisa.heimbauer@gmail.com), Michael J. Beran, and Michael J. Owren (Georgia State Univ., Atlanta, GA 30302-5010)

The ability of human listeners to understand speech even in altered or synthetic forms is argued to be evidence of uniquely human processing abilities. However, extensive early experience with speech may also contribute to this capability. To investigate this issue two experiments were designed to test the ability of Panzee, a 22-year-old language-trained chimpanzee, to recognize words in synthetic form. Like a human child, she was reared from infancy by human caregivers who routinely spoke to her. She communicates with humans by identifying words using graphical symbols (lexigrams). Experimental training and testing were conducted with two different sets of 24 familiar words presented one-per-trial in natural, voiced-only, or noise-vocoded forms, with Panzee choosing one of four lexigrams presented on a computer monitor. Experiment 1 showed equivalent performance with words heard in natural form versus voiced-only versions, resynthesized from only voiced components of a word. Noise-vocoded words presented in Experiment 2 simulated effects of hearing using a cochlear implant and were based on amplitude-modulated noise bands. Performance with these sounds was significantly higher than chance, but also lower than with natural words. Results suggest specialized processing mechanisms are not necessary to speech perception in the absence of traditional acoustic cues.

3pSC17. Identifying the common problems of English-to-Japanese consecutive interpretations performed by Japanese intepreting students. Kinuko Takahashi (Sophia Univ. #306, 4-1-10, Dai, Kamakura-city 247-0061 Japan, kinuko@dc4.so-net.ne.jp)

The purpose of the present paper is to identify the problems that Japanese interpreting students may encounter and to establish causes for them in the expectation that they may help devise a training method. It has been long wondered why some students become successful interpreters and others do not. In order to answer this question, the research was launched to identify the problems that interpreting students encountered. For this purpose, nine

interpreting students were asked to consecutively interpret four different English texts. It was discovered that the participants had a tendency for omissions of interpreting, and omissions occurred due to a number of causes. One of them was a problem of speech perception. One of the participants, for example, was not able to distinguish "breeze" from "bleed." To one of the participants, "attendance" sounded like "tendence." These problems seemed to require the participants to take more time to interpret. As a result, the participants failed to listen to the following parts, eventually making further omissions. Therefore, it is necessary for the participants to learn "top-down" processing of comprehension. However, it is necessary to examine the relation between the level of interpreting skills and speech perception ability.

3pSC18. Perceptual normalization for variation in speaking style. Antonia D. Vitela, Sarah C. Sullivan, and Andrew J. Lotto (Speech, Lang. and Hearing Sci., Univ. of Arizona, 1131 E. 2nd St., Tucson, AZ 85721, adv1@email.arizona.edu)

Ladefoged and Broadbent [J. Acoust. Soc. Am. **29**, 98–104 (1957)] demonstrated that listeners will shift their categorization of a target vowel as a function of acoustic characteristics of a preceding carrier phrase. These results have been interpreted as an example of perceptual normalization for variability resulting from differences in talker anatomy. The present study examined whether listeners would normalize for acoustic variability resulting from differences in speaking style within a single talker. Two vowel series were synthesized that varied between central and peripheral vowels (the vowels in "beat"-"bit" and "bod"-"bud"). Each member of the series was appended to one of four carrier phrases that were spoken in either a "clear" or "reduced" speech style. Participants categorized vowels in these eight contexts. A reliable shift in categorization as a function of speaking style was obtained for three of four phrase sets. This demonstrates that phrase context effects can be obtained with a single talker. However, the directions of the obtained shifts are not reliably predicted on the basis of the speaking style of the talker. Instead, it appears that the effect is determined by an interaction of the average spectrum of the phrase with the target vowel. [Work supported by NIH-NIDCD.]

3pSC19. Neural dynamics of speech perception: Phonemic restoration in noise using subsequent context. Sohrob Kazerounian and Stephen Grossberg (Dept. of Cognit. and Neural Systems, Boston Univ., 677 Beacon St., Boston, MA, 02215, sohrob@cns.bu.edu)

Phonemic restoration describes a class of auditory illusions during which broadband noise replacing a deleted phoneme can cause the perceptual restoration of the deleted phoneme. Phonemic restoration exemplifies the brain's ability to complete and understand speech and language in noisy environments. It also clarifies how both past and future acoustical events can contextually guide completion of a percept that is occluded by noise, and highlights that conscious speech is due to bottom-up and top-down contextual interactions that can operate across hundreds of milliseconds. This work develops a neural model that quantitatively simulates restoration phenomena, including the forward development in time of a conscious speech percept even in cases where future events control how previously presented acoustic events are heard. The model clarifies how acoustic items are stored in short-term working memory, and how they interact reciprocally with unitized representations of item sequences, or list chunks, to generate a resonant wave of activation that embodies the consciously heard percept. Model simulations clarify why the presence of noise is necessary for restoration to occur, and why in the absence of noise a silence gap is perceived. These properties are traced to the brain's ability to rapidly and stably learn language.

3pSC20. Towards produtive language mapping: A magnetoencephalography study of letter naming. Yang Zhang (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, Minneapolis, MN 55455), Wenbo Zhang, and Joel Landsteiner (United Hospital, St. Paul, MN 55102)

This study aims to examine the sensitivity and specificity of magnetoencephalography (MEG) integrated with magnetic resonance imaging (MRI) for productive language mapping using a simple letter naming task. Six healthy adult volunteers participated after screening for handedness and medical history of speech, language, hearing, and vision. The data were collected using a 148-channel whole-head MEG system (Magnes WH2500, 4D Neuroimaging, San Diego) at the Magnetic Source Imaging Laboratory in United Hospital, St. Paul. During the recording session in a magnetically shielded room, the subjects were instructed to pay attention to the visual stimuli and articulate the letter from the English alphabet as soon as it was shown on the screen. The averaged epochs and trial-to-trial raw data were analyzed using Brain Electrical Source Analysis (BESA) and BrainVoyager software. Although visual processing of the letters showed similar temporal and spatial activation patterns, letter naming did not show uniform activation patterns in the Broca's area. Large inter-subject variability was found in terms of superior temporal, inferior parietal, motor, and supplemental motor involvement. The results are discussed with respect to the technical challenges in implementing MEG with millisecond resolution for presurgical productive language mapping and potential risks of sensorimotor and language deficits.

3pSC21. An investigation of brain activation patterns in response to speechreading. Edward T. Auer, Jr. (Dept. of Speech-Lang.-Hearing, Univ. of Kansas, 1000 Sunnyside Ave., Rm. 3001, Lawrence, KS 66045, auer@ku.edu)

The ease of recognizing spoken words on the basis of optical input alone, or speechreading, is known to systematically vary as a function of the properties of the stimulus materials. The current study investigated the relationship between cortical responses and the intelligibility of the stimulus materials. Open-set identification performance outside the magnet was used to directly measure the intelligibility of the stimulus materials presented in the magnet as well as participant speechreading ability. fMRI was used to measure cortical response to viewing spoken word stimuli that were predicted to vary in their ease of identification. Videos of spoken words were contrasted with still face images in a block-design fMRI experiment. Blocks containing sets of words predicted to be easy speechread were contrasted with blocks predicted to be difficult to speechread. Preliminary analyses of cortical response to speechreading are consistent with previous studies [Hall *et al.*, J. Cog. Neuro. **17**(6), 939–953 (2005)]. Furthermore, evidence of increased activation in BA 37 and BA 44 for easy words versus hard words was obtained. These results are consistent with the hypothesis that specific areas are sensitive to the intelligibility of speech read stimuli. [Work supported by a grant from NIH/NIDCD (R01DC04856).]

3pSC22. Domain-specific processing of Mandarin tone. Yue Wang (Dept. of Linguist., Simon Fraser Univ., Burnaby, BC, Canada, V5A 1S6), Dawm Behne (Norwegian Univ. of Sci. and Technol., Trondheim, Norway), Angela Cooper, and Jung-yueh Tu (Simon Fraser Univ., Burnaby, BC, V5A 1S6 Canada)

Lexical tone has generally been found to be processed predominantly in the left hemisphere. However, given that tone is carried by a syllable or a word with segmental information and distinctive meaning, the processing of tone may not be easily disentangled from that of the phonetic segments and word meaning [P. Wong, Brain Res. Bull. **59**, 83–95 (2002)]. Indeed, previous research has not examined the lateralization of tone independent of segmental and lexical semantic information. The present study explores how syllable-based tonal processing in Mandarin Chinese interacts with these different linguistic domains. Using dichotic listening, native Mandarin participants were presented with monosyllabic tonal stimuli constructed with the following different linguistic attributes: (1) real Mandarin words with tonal, segmental phonetic, and lexical semantic information; (2) Mandarin nonwords with tonal and segmental, but no semantic information; (3) nonwords with non-Mandarin segments (i.e., no native segmental or semantic information); and (4) hums of tones (acoustic pitch information) without any segmental or semantic components. Results from these conditions show significant differences in lateralization patterns and are discussed in terms of the integration of acoustic as well as pre- and post-lexical linguistic domains in lexical tone processing. [Work supported by NSERC.]

3pSC23. Mismatch negativity reflects the affect of training on speech perception. McNeel Gordon Jantzen and Katherine Cella (Dept. of Psych., Western Washington Univ., 516 High St., Bellingham, WA 98225)

An adult's ability to perceive speech sounds that do not occur in their native language is influenced by the relationship between the new and native language phonology and by the individual's capacity to perceive acoustic differences. Individual differences are important: even after phonetic train-

ing, adults with similar language backgrounds exhibit significant individual differences in their ability to learn difficult non-native speech sounds [Polka (1991); Pruitt, Strange, Polka, & Aguilar (1990)]. Neurophysiologic changes due to learning are also reflected in the mismatch negativity (MMN) response and are indicative of the effects of training [Naatanen *et al.* (1993) and Kraus *et al.* (1995)]. The present experiment included two groups of six normal-hearing monolingual American-English speakers. The experimental group was trained to identify a voiced, unaspirated, dental stop consonant. The control group received no training on the dental. Electrophysiologic responses were measured for both groups in response to the dental and alveolar contrasts using a perceptual mapping procedure. Training improved the experimental group's ability to perceive the non-native contrast. Training effects were also reflected in the MMN response, as observed by increased duration and decreased onset latency.

3pSC24. The interaction of location with acoustic scale in concurrent speech recognition. David T. Ives, Martin D. Vestergaard, and Roy D. Patterson (Dept. of Physio., Development and Neurosci., Univ. of Cambridge, Downing St., Cambridge CB2 3EG, UK)

Location and acoustic scale cues have been shown to have a significant effect on the recognition of speech in multispeaker environments. The interaction of these two cues is less well understood. In this study, subjects are presented with two triplets of concurrent speech syllables with similar temporal envelopes, and asked to recognize a specific target syllable. The task was made more or less difficult by changing the location of the distracting speaker, the scale difference between the two speakers and the relative level of the two speakers. Scale differences were produced by changing the vocal tract length and glottal pulse rate of resynthesized speech: 32 scale differences were used. Location cues were produced by convolving heat-related transfer functions with the stimulus. The target speaker was located directly to the front of the listener and the distracting/masking speaker located from one of five locations (0, 4, 8, 16, 32 deg) on the 0 deg horizontal plane. Target-to-masker ratios of 0 and −6 dB were used. The results show that direction and scale differences cues do interact and this interaction is greatest when directional and speaker scale cues are small. [Research supported by the U.K. Medical Research Council (G0500221, G9900369).]

3pSC25. Does training improve consistency of roughness judgments in a matching task? David A. Eddins (Dept. of Otolaryngol., Univ. of Rochester, 2365 S. Clinton Ave., Rochester, NY 14618, David_Eddins@URMC.Rochester.edu), Rahul Shrivastav, and Sona A. Patel (Univ. of Florida, Gainesville, FL 32611)

Shrivastav *et al.* [J. Acoust. Soc. Am. **119**(5), 3340 (2006)] reported a matching task in which a square wave modulated sawtooth wave (signal) was matched to a series of vowels (standards) to estimate the magnitude of roughness in voices. Results suggested that listeners found it difficult to isolate roughness from other voice quality dimensions such as breathiness. In the present experiment, a brief training session was added prior to the matching task to ensure that listeners were attending to variations in roughness alone. This training consisted of a rank-ordering task in which listeners sorted the 34 test samples of the vowel /a/ in ascending order of roughness. For feedback, these rankings were compared to the rankings made by expert listeners. This criterion was used to ensure that all participants understood the dimension of voice quality under study. Listeners who met a specific eligibility criterion completed a modulation matching task to familiarize them with the matching task itself. Finally, thresholds for the roughness matching task were obtained and compared to thresholds obtained without training. The extent to which task-related training can help listeners make perceptual judgments for specific dimensions of voice quality will be discussed.

3pSC26. Masking release at low sensation levels. Peggy Nelson, Elizabeth Anderson Crump, Yingjiu Nie, and Michelle Hawkinson-Lewis (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, peggynelson@umn.edu)

Previous results have shown that listeners with sensorineural hearing loss (SNHL) obtain about half of the masking release of their normal-hearing (NH) counterparts. When speech is amplified sufficiently, listeners with SNHL may score like NH listeners in quiet and in steady noise, yet may obtain only half of the expected release from gated noise. We hypothesize that some of that deficiency may occur because of the impaired listeners' low speech sensation levels, which results in decreased usefulness of the speech signal in the noise dips. In the current study, NH listeners were tested for their recognition of IEEE sentences in quiet, in steady noise, and in gated noise with the speech presented at varying sensation levels. At low levels (10–15 dB SL), NH listeners scored nearly 100% correct in quiet. In steady noise (at −10 dB signal-to-noise ratio) scores for low-level stimuli were also similar to those obtained at higher SLs. However, at low SLs in gated noise, NH listeners demonstrated less masking release than expected, suggesting that audibility of speech in the dips of noise is important for masking release even when performance in quiet and in steady noise seems optimized. [Work supported by NIH 5R01DC008306.]

3pSC27. Listening to natural versus cell phone speech on multiple simultaneous tasks. Srinivasan Nirmal Kumar (Dept. of Special Educ. and Commun. Disord., Barkley Memorial Ctr., Univ. of Nebraska-Lincoln, Lincoln, NE 68583) and Carrell Thomas (Univ. of Nebraska-Lincoln, Lincoln, NE 68583)

In typical listening environments, attention is often divided and may have different effects on automatic and controlled processes. Automatic processing is a fast, parallel process not limited by short-term memory, requires little subject effort, but requires extensive consistent training to develop. Controlled processing is a comparatively slow, serial process limited by short-term memory, requires subject effort, and little training to develop [Schneider and Shiffrin (1977)]. In the present study a methodology was developed to examine effects of controlled and automatic distracters on the perception of distorted speech. Specifically, perception of natural and cell phone speech was investigated while listeners performed simultaneous visual and motor tasks. Young, normal-hearing native speakers of English were presented with SPIN sentences [Kalikow *et al.* (1977)] in a background of multitalker babble [Bilger *et al.* (1984)] using natural and cell phone speech. Prior to the start of the experiment, participants had been trained on visual task using either consistent mapping technique or varied mapping technique. Word recognition scores, pursuit rotor performance, and visual task performance were compared for natural and cell phone speech. The relationship between consistently mapped and variably mapped distracters on perceptual and behavioral performance provides information necessary for more detailed models relevant to real-world environments.

3pSC28. Seeing a speaker's face helps stream segregation for younger and elderly adults. Alexandra Jesse (Max Planck Inst. for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands, Alexandra.Jesse@mpi.nl) and Esther Janse (Utrecht Inst. of Linguist. OTS and Max Planck Inst. for Psycholinguistics, Nijmegen, The Netherlands)

Listening to a speaker while hearing one or more competing speakers in the background can be a challenging task, especially for elderly adults. We show that younger and elderly listeners (above the age of 65) with varying degrees of age-related hearing loss benefit in stream segregation and speech processing from seeing the target speaker talk in addition to hearing an audio mix of a target and a competing speaker. This audiovisual benefit was found for response accuracy and speed in a phoneme monitoring task, where participants indicated by button press the occurrence of given target phonemes in the monitored speech of the target speaker. The audiovisual benefit was larger for younger than for older adults, despite their equivalent performance on auditory-only trials and in an off-line phoneme lip-reading task. Better lip-reading performance, however, predicted a larger audiovisual benefit. The audiovisual benefit was found for both highly visible phonemes (/p/) and poorly visible phonemes (/k/), but was modulated in its size by segmental visibility. The audiovisual benefit therefore arises from local segmental visual information but is also at least partially driven by audiovisual synchrony information that aids in attending to the target speaker.

3pSC29. The eyes' footprints on the ears: An investigation of short-term speech intelligibility change. Jing Liang (Dept. of Psych. and Ctr. for Cognit. and Social Neurosci., Univ. of Chicago, 5848 S Univ. Ave., Chicago, IL 60637, liang@uchicago.edu), Steven L. Small, and Howard C. Nusbaum (Univ. of Chicago, Chicago, IL)

Although observation of mouth movements improves auditory speech perception, the extent to which visual information aids speech perceptual learning and affects subsequent audio-only speech perception remains

unknown. The current study investigates whether visual speech information, specifically, the synchronized mouth movements of the talking face during training, helps the listener to perceive audio-only speech more effectively after training. An experimental group was trained to recognize audiovisual words presented in noise, while a control group was trained on the same audio speech signal in noise, but with no accompanying mouth movements. The control group provides a baseline estimate of what listeners can learn from the speech signal alone without visual information from the talker's mouth movements. Both groups were tested on audio-only speech in noise before and after training. All novel words were used in the pretest, training, and posttest. The results demonstrate that visual information aids learning to recognize audio-only speech in noise, suggesting visual information from a talker's mouth movements during training might change auditory coding of acoustic speech signals.

3pSC30. Effects of audio-visual speech information on recognition memory of spoken words. Kauyumari Sanchez, Rachel M. Miller, and Lawrence D. Rosenblum (Dept. of Psych., Univ. of California, Riverside, 900 University Ave., Riverside, CA 92521)

Audio-visual speech has generally been found to contain more usable information than audio-only speech. However, there is conflicting evidence of whether seeing the face of a speaker facilitates memory for spoken words [e.g., Sheffert *et al.*, Cog. Tech. **8**, 42–50 (2003)]. To address this issue, three experiments examined whether an audio-visual benefit would be observed on a word recognition task. Experiment 1 compared recognition of spoken words both presented and tested in audio-visual versus audio-only forms. Audio-visual word stimuli were recognized significantly better than audio-only words. Experiment 2 tested whether this benefit was due to the presence of visible articulatory information or simply more information in general. Recognition of words presented (and tested) in audio-only, audio-visual, and audio with accompanying static face image conditions were compared. Words presented in audio-visual (dynamic face) form were recognized better than audio-only and audio-static face stimuli. To test whether the benefit of audio-visual presentation was due to encoding, retrieval, or both, Experiment 3 crossed presentation modality (audio-only versus audio-visual) from presentation to test phases. Results showed that the audio-visual form is required in both encoding and retrieval to be beneficial.

3pSC31. Influence of visual speech information on the identification of foreign accented speech. Rebecca K. Reed and Edward T. Auer, Jr (Dept. of Speech-Lang.-Hearing, Univ. of Kansas, 1000 Sunnyside Ave., Rm. 3001, Lawrence, KS 66045, rxxr@ku.edu)

Seeing a native talker's face improves speech intelligibility in noise for native perceivers. The intelligibility of foreign accented speech (English spoken by native talkers of Mandarin) is more susceptible to the effects of noise than the speech of native talkers. [Rogers *et al.*, Lang Speech 47(2), 139–154 (2004)]. The current experiment investigated the influence of seeing the non-native talker's face on the intelligibility of speech presented in noise. Ten talkers (nine non-native and one native) were recorded producing 155 sentences each. Talker-specific speech shaped noise was mixed with the audio of the sentences. Eighty-one native perceivers of English (nine per non-native talker) responded to 20 sentences spoken by the native talker and 28 by the non-native talker in each condition (audio-alone, visual-alone, audiovisual). Open set identification responses were scored as percent words correct for each condition and talker. Preliminary analyses demonstrate that although seeing the non-native talker increases intelligibility, the gain is significantly reduced compared to the native talker. Results will be discussed in terms of potential sources of the audiovisual gain. These sources are hypothesized to be differentially sensitive to the distortions arising in non-native speech. [Work supported in part by NIH/NIDCD DC04856 and the University of Kansas, UGRA.]

Session 3pUW

Underwater Acoustics and Acoustical Oceanography: Physics-Based Undersea Clutter Model Verification and Validation II

Juan I. Arvelo, Cochair
Applied Physics Lab., Johns Hopkins Univ., Laurel, MD 20723-6099

Kenneth G. Foote, Cochair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Timothy K. Stanton, Cochair
Dept. of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Contributed Paper

1:00

3pUW1. Propagation focusing in the context of clutter statistics. Chris Harrison (NURC, Viale S. Bartolomeo, 400, 19126 La Spezia, Italy, harrison @nurc.nato.int)

Fluctuations in reverberation (clutter) from scatterers at long range in shallow water can be affected by many mechanisms. Their statistics, in particular the scintillation index, depend not only on the statistics of the scatterers (their physical distribution, shape, orientation, etc.) but on the outward and return propagation paths. Focusing effects due to forward scattering from weak undulations in the seabed will be presented. Existing Fresnel–Kirchhoff theory for monostatic, vertical incidence on a two-dimensional surface is adapted to demonstrate the behavior of the intensity's scintillation index in bistatic geometry. In addition, using a simulation of bistatic geometry and a one-dimensional surface, the intensity statistics are extended to include autocorrelation function and probability distribution. Bistatic geometries include source and receiver moving with constant separation, and moving source with fixed receiver. Fairly modest vertical undulations can result in significant fluctuation due to reflection focusing with scintillation indices possibly greater than unity.

Invited Paper

1:15

3pUW2. Estimating higher moments of shallow water reverbation for non-Gaussian scatterer distributions. Kevin D. LePage (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19126 La Spezia, Italy)

Reverberation is described as a random process by which deterministic propagation components weight and sum scattering amplitudes over the sonar footprint. A method of estimating the second, third, and fourth monents of the scattered field pressure is described for scatterers which have a Chi Squared model for the amplitude pdf. Comparison of theoretical estimates of the fourth moment to Monte Carlo estimates of the same quantity obtained using a NRL 2-way parabolic equation PERM show good agreement. [Work supported by ONR.]

Contributed Papers

1:35

3pUW3. Measurement and modeling of broadband Bragg scattering from a sinusoidal surface. Dajun Tang and Darrell R. Jackson (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, djtang @apl.washington.edu)

While the mechanism of Bragg scattering is well known, most experimental work has been concentrated in the area of narrow band sound sources and in the far-field. Motivated by underwater detection problems in the presence of sediment ripple fields, we report laboratory measurements of broadband sound scattering from a sinusoidal surface machined on a polyurethane board. The surface has a wavelength of 8 mm and peak-to-peak height of 2 mm. Coherently scattered sound data were taken in near-field geometries and in the frequency band of 150–400 kHz. The measurement geometry is such that a broad range of Bragg angles corresponding to the frequency band are covered. We observe that the scattered sound demonstrates a down chirp time dependence when the incident sound is a short pulse. Models based on first order perturbation theory were developed which explain the observed scattered sound in both magnitude and phase. In addition, we also measured second order Bragg scattering. This motivates modeling efforts on higher order Bragg scatter. [Work supported by ONR.]

1:50

3pUW4. Validation of a physical model for surface clutter. Trudy L. Philip, Bruce K. Newhall, and Juan I. Arvelo (Johns Hopkins Univ. Appl. Phys. Lab., 11100 Johns Hopkins Rd., MS 8-220, Laurel, MD 20723)

There is a strong need to more accurately represent active sonar system false contacts in various environments and conditions for the purpose of active acoustic simulation and synthetic training. A computationally viable approach for the generation of physics-based false contacts in a raw beamformed time series that can be injected into a sonar processor was previously developed for bottom clutter. That model is now extended for clutter due to rough ocean surface scattering, and is compared with ASIAEX01 data. Considerations in extending the existing model include: characterizing appropriate statistics of the rough surface, adding surface scattering clutter to the sonar model, evaluating the realism of the simulation, and evaluating the computational burden (for real-time trainers). The output of the simulation was analyzed for those considerations and then compared to the data. The use of a vertical line array allowed physical mechanisms in the data to be isolated, and their clutter data statistics separately determined. After comparing data statistics to predictions, the need for additional modeled clutter mechanisms was assessed. This effort was conducted under the auspices of the Undersea Warfare Business Area Independent Research and Development program of the Johns Hopkins University Applied Physics Laboratory.

Plenary Session and Awards Ceremony

Mark F. Hamilton, Chair
President, Acoustical Society of America

Business Meeting of the Acoustical Society of America

Presentation of Certificates to New Fellows

Russell E. Berger, II
Suzanne E. Boyce
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

Presentation of Acoustical Society Awards

Medwin Prize in Acoustical Oceanography to Martin Siderius

R. Bruce Lindsay Award to Kelly J. Benoit Bird

2008 Silver Medal in Speech Communication to Winifred Strange

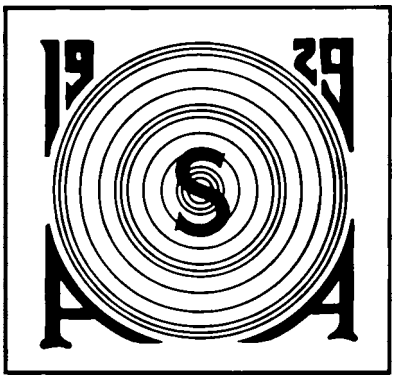
Gold Medal to Thomas D. Rossing

Presentation of Vice President's Gavel

Presentation of President's Tuning Fork

ACOUSTICAL SOCIETY OF AMERICA

R. BRUCE LINDSAY AWARD



Kelly J. Benoit-Bird

2009

The R. Bruce Lindsay Award (formerly the Biennial Award) is presented in the Spring to a member of the Society who is under 35 years of age on 1 January of the year of the Award and who, during a period of two or more years immediately preceding the award, has been active in the affairs of the Society and has contributed substantially, through published papers, to the advancement of theoretical or applied acoustics, or both. The award was presented biennially until 1986. It is now an annual award.

PREVIOUS RECIPIENTS

Richard H. Bolt	1942	Ilene J. Busch-Vishniac	1987
Leo L. Beranek	1944	Gilles A. Daigle	1988
Vincent Salmon	1946	Mark F. Hamilton	1989
Isadore Rudnick	1948	Thomas J. Hofler	1990
J. C. R. Licklider	1950	Yves H. Berthelot	1991
Osman K. Mawardi	1952	Joseph M. Cuschieri	1991
Uno Ingard	1954	Anthony A. Atchley	1992
Ernest Yeager	1956	Michael D. Collins	1993
Ira J. Hirsh	1956	Robert P. Carlyon	1994
Bruce P. Bogert	1958	Beverly A. Wright	1995
Ira Dyer	1960	Victor W. Sparrow	1996
Alan Powell	1962	D. Keith Wilson	1997
Tony F. W. Embleton	1964	Robert L. Clark	1998
David M. Green	1966	Paul E. Barbone	1999
Emmanuel P. Papadakis	1968	Robin O. Cleveland	2000
Logan E. Hargrove	1970	Andrew J. Oxenham	2001
Robert D. Finch	1972	James J. Finneran	2002
Lawrence R. Rabiner	1974	Thomas J. Royston	2002
Robert E. Apfel	1976	Dani Byrd	2003
Henry E. Bass	1978	Michael R. Bailey	2004
Peter H. Rogers	1980	Lily M. Wang	2005
Ralph N. Baer	1982	Purnima Ratilal	2006
Peter N. Mikhalevsky	1984	Dorian S. Houser	2007
William E. Cooper	1986	Tyrone M. Porter	2008

CITATION FOR KELLY J. BENOIT-BIRD

. . . for contributions in marine ecological acoustics

PORTLAND, OREGON• 20 MAY 2009

Kelly J. Benoit-Bird graduated from Maloney High School in Meriden, Connecticut where she was a member of the National Honor Society and was voted most likely to succeed by her classmates. She obtained a bachelor's degree with honors in 1998 from Brown University majoring in Aquatic Biology. In the fall of 1998, she enrolled at the University of Hawaii to work toward a Ph.D. in Zoology. As a graduate student, Kelly performed research on the properties of the mesopelagic boundary community, the foraging behavior of spinner dolphins, and the acoustic properties of deep dwelling Hawaiian lutjanid snappers. She earned a Ph.D. in 2003 and continued as a post doctoral fellow at the Hawaii Institute of Marine Biology, University of Hawaii, for another year.

Kelly's research while in Hawaii has involved the use of various acoustic techniques to study the ecology of the coastal ocean environment, including the sound scattering layer of micronekton in the mesopelagic boundary community around the Hawaiian Islands and the interaction of these organisms with higher trophic level predators, including spinner dolphins. She also participated in research to determine the acoustic backscatter characteristics of deep dwelling lujanid snappers, some of which are severely threatened in the main Hawaiian Islands.

In July of 2004, Kelly became an assistant professor in the College of Oceanic and Atmospheric Sciences at Oregon State University. Since that appointment, she has distinguished herself by her innovative application of multi-frequency acoustics to the investigation of the predator-prey dynamics of fish and zooplankton within planktonic thin layers. In another recent project, she demonstrated that the biomass and behavior of soft-bodied squid in the Gulf of California can be quantified with appropriate application of acoustics. She continues to push for new approaches to understanding ecological interactions across a range of spatial scales, using multibeam acoustics on autonomous underwater vehicles and well as from small vessels.

Kelly is truly one of the pioneers in the use of acoustics to address ecological issues in the marine environment. She already has 27 publications in peer reviewed journals in her young career. As a result of her skill and hard work, Kelly has made substantial contributions to the understanding of coastal ecological systems. Her contributions, together with a number of collaborators, have been reported for the first time in scientific journals. They include migration behavior of micronekton in the mesopelagic boundary community (MBC), acoustic properties of MBC organisms, interaction of spinner dolphins and MBC prey and sonar characteristics of lujanid snappers. The migration behavior was first reported in *Marine Ecology Progress Series* in 2001; acoustic properties of MBC organisms in the *Journal of the Acoustical Society of America (JASA)*, also in 2001; interaction of spinner dolphins and prey in *Behavioral Ecology and Sociobiology* in 2003 and *JASA* in 2009; and the sonar characteristics of lujanid snappers in two papers in *JASA* in 2003.

Kelly, by her outstanding use of graphics in displaying her experimental data and explaining her results, is perhaps the singular most influential individual contributing to a significant improvement in the quality of presentations in Acoustic Oceanography and Animal Bioacoustics sessions. Between the years 2000 and 2003, she won many best student paper awards at ASA meetings. It is a great accomplishment to win just one "Best Student Paper" award but to win two awards at same meeting is a unique achievement. Kelly accomplished this task not once, but twice. She also won another best student paper award in Acoustical Oceanography and she placed 2nd in Acoustical Oceanography at the First Pan-American/Iberian Meeting on Acoustics, in Cancun, Mexico. The total of five best student paper awards and one 2nd place finish is not only unheard of, but is hardly imaginable. What Kelly has achieved is absolutely unprecedented in the annals of the ASA. She was also one of the nine young presenters at the ASA's 75th Anniversary Celebration Plenary Session in New York in 2004.

Kelly's achievements have not been confined to ASA meetings. She won the Excellence in Communications Award at the 15th Biennial Conference on the Biology of Marine Mammals (by a non-student) in 2003 and the Best Young Presenter Award at the European Congress on Underwater Acoustics in 2004. Her outstanding research has won

her other awards, which include the Presidential Early Career Award for Scientists and Engineers FY 2005 (PECASE) in 2006, the Kavli Frontiers Fellowship from the National Academy of Sciences in 2007, and the Ocean Sciences Early Career Award of the American Geophysical Union in 2008.

We cannot think of anyone more deserving of the Bruce Lindsay award than Kelly Benoit-Bird.

WHITLOW W.L. AU
TIMOTHY J. COWLES

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal in

Speech Communication

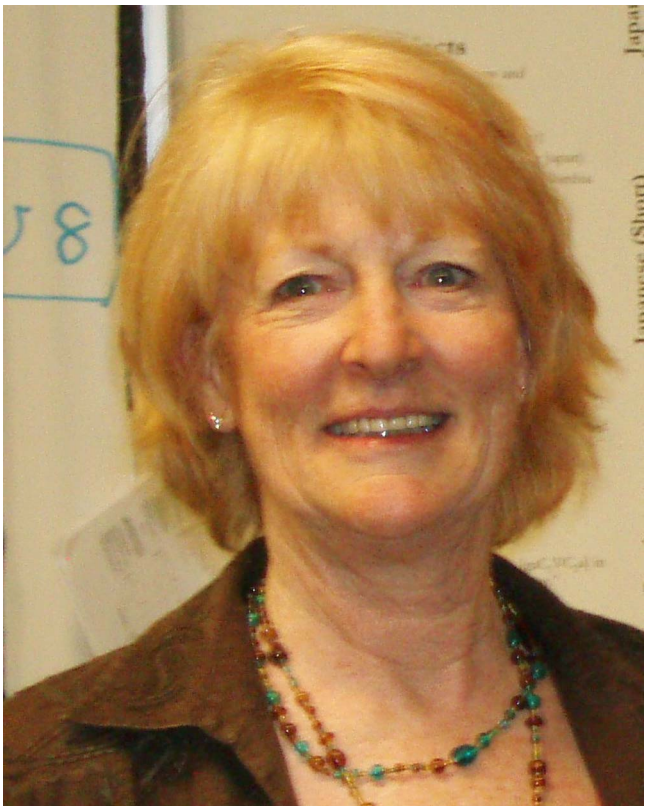
Winifred Strange

2008

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

Frank S. Cooper	1975	Peter Ladefoged	1994
Gunnar Fant	1980	Patricia K. Kuhl	1997
Kenneth N. Stevens	1983	Katherine S. Harris	2005
Dennis H. Klatt	1987	Ingo R. Titze	2007
Arthur S. House	1991		



CITATION FOR WINIFRED STRANGE

. . . for contributions to understanding speech preception.

PORTLAND, OREGON • 20 MAY 2009

Winifred Strange was born in Red Bluff, California, the younger of two girls born to Verdon Earl Strange and Winifred Rose Strange. Winifred is known by many as 'Pinky,' a nickname coined when her older sister commented on her rosy newborn appearance, and a name that persisted due to the flaming red hair and spirited personality that soon emerged. After graduating from Chico State College in California, Winifred became a graduate student in the Department of Psychology in the early 1970s, where she was acknowledged as one of the University of Minnesota's best teaching assistants. Pinky was the teaching assistant for an intense, yearlong graduate seminar taught by Professor James J. Jenkins, himself a legend at the University of Minnesota. An interdisciplinary group of graduate students interested in the new field of 'cognitive science' adopted the seminar's goal, which was to dissect, sentence-by-sentence, one of the most exciting new theories in cognitive psychology, described in MIT Professor Noam Chomsky's book *Syntactic Structures*. Pinky's highly spirited repartees with the graduate students convinced many of them to alter their careers and devote their lives to this field.

Winifred later married Jenkins, and became step-grandmother to 7, step-great grandmother to 5, and spiritual grandmother to one Christopher Chisolm, son of Richard Chisolm and Teresa Hnath-Chisolm. At the University of South Florida, where Winifred and Jim took faculty positions in 1982, they mentored a new generation of students. In 1998, Winifred took a faculty position at the City University of New York–Graduate School and University Center, where she continues today to play an important role in mentoring doctoral students.

Winifred's scientific contributions are in the area of adult speech perception. She was a pioneer in research on cross-language speech perception, in which she focused on the effects of perceptual training and nonnative language experience. Her work on vowel perception changed our notions of the extent to which speech sounds are processed using dynamic as opposed to static acoustic information. Most experiments conducted in the first 30 years of research on speech perception used stimuli that were generated by computer and contained steady-state formant frequencies. In experiments published in the *Journal of the Acoustical Society of America* beginning in the late 1980s, Strange showed that one could remove the steady-state centers of vowels, leaving only the formant transitions into and out of the vowel, and perception of the vowel remained excellent. These findings on "vowel-less" syllables were surprising, and were initially met with skepticism. However, further research produced a body of evidence that weakened alternative interpretations and led to the acceptance of the idea that the dynamic portions of the onsets and offsets of coarticulated vowels contributed importantly to vowel perception. From that point on, researchers had to take into account the fact that listeners were attuned to the dynamic properties of speech sounds, not just their static steady-state values.

One of the biggest challenges of this work is Winifred's continuing commitment to understanding how vowels are perceived in natural co-articulated syllables. Research taking the dynamic specification view in other languages is beginning to show how vowel spectral dynamics and their perception vary across languages. Recent work compares the influence of phonetic context, prosodic context, and speech style on the acoustic characteristics and the cross-language perception of languages such as English, German, Japanese, and French. Her detailed studies of how acoustic similarity relates to perceptual similarity has contributed importantly to an in-depth, and ecologically more valid, understanding of cross-language perception of vowels and of the language-specific influences of consonantal context on acoustic vowel characteristics.

Winifred Strange has contributed greatly to the Acoustical Society of America (ASA) as a past Associate Editor for Speech for the *Journal of the Acoustical Society of America*, as a Member of the Executive Council from 2001-04, and as an active member of the Speech Technical Committee. Winifred has worked behind the scenes on numerous projects, including the ASA's 75th Anniversary Celebration Committee. Many students and junior colleagues have been exposed to her contagious enthusiasm for our field, and

she continues to introduce new students to the ASA. Awarding the Silver Medal in Speech Communication to Winifred Strange recognizes her contributions to the science of speech communication as well as her many contributions to the Acoustical Society of America.

LINDA POLKA
OCKE-SCHWEN BOHN
PATRICIA KUHL

GOLD MEDAL
of the
Acoustical Society of America

Thomas D. Rossing

2009

The Gold Medal is presented in the spring to a member of the Society, without age limitation, for contributions to acoustics. The first Gold Medal was presented in 1954 on the occasion of the Society's Twenty-Fifth Anniversary Celebration and biennially until 1981. It is now an annual award.

PREVIOUS RECIPIENTS

Wallace Waterfall	1954	Richard K. Cook	1988
Floyd A. Firestone	1955	Lothar W. Cremer	1989
Harvey Fletcher	1957	Eugen J. Skudrzyk	1990
Edward C. Wente	1959	Manfred R. Schroeder	1991
Georg von Békésy	1961	Ira J. Hirsh	1992
R. Bruce Lindsay	1963	David T. Blackstock	1993
Hallowell Davis	1965	David M. Green	1994
Vern O. Knudsen	1967	Kenneth N. Stevens	1995
Frederick V. Hunt	1969	Ira Dyer	1996
Warren P. Mason	1971	K. Uno Ingard	1997
Philip M. Morse	1973	Floyd Dunn	1998
Leo L. Beranek	1975	Henning E. von Gierke	1999
Raymond W. B. Stephens	1977	Murray Strasberg	2000
Richard H. Bolt	1979	Herman Medwin	2001
Harry F. Olson	1981	Robert E. Apfel	2002
Isadore Rudnick	1982	Tony F. W. Embleton	2002
Martin Greenspan	1983	Richard H. Lyon	2003
Robert T. Beyer	1984	Chester M. McKinney	2004
Laurence Batchelder	1985	Allan D. Pierce	2005
James L. Flanagan	1986	James E. West	2006
Cyril M. Harris	1987	Katherine S. Harris	2007
Arthur H. Benade	1988	Patricia K. Kuhl	2008

CITATION FOR THOMAS D. ROSSING

. . . for contributions to musical acoustics, leadership in science education, and service to the Society.

PORTLAND, OREGON • 20 MAY 2009

On rare occasions one encounters an individual, who combines excellence in a somewhat narrow field, and depth in many broad areas, with incisive intellectual curiosity, and a rare ability to share insights. Such a remarkable personality is Tom Rossing. Born in Madison, South Dakota, Tom received an undergraduate degree in physics and mathematics from Luther College, and went on to obtain MS and PhD degrees in Physics from Iowa State University. After three years in industry he spent 14 years on the physics faculty of St. Olaf College, six of those as department chair. As was the case for many of us Tom returned to his second love: Music. In the early 70's Tom established an acoustics research program at Northern Illinois University specializing in musical acoustics. That program became a center of excellence, a source of inspiration for innumerable students, a springboard for countless young scientists entering the field, and a resource for scientists in musical acoustics world-wide. Since 2004 he has been a visiting professor in the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University.

Tom quickly established his reputation with numerous publications on percussion instruments. Other research publications were followed by two extremely influential works: *The Science of Sound*, one of the most frequently used texts for a general science course aimed at music majors, and *The Physics of Musical Instruments,* co-authored with Neville Fletcher, an absolute must for anyone entering the field of musical instrument research. The volume on *Acoustics of Bells* in the Benchmark Papers in Acoustics Series was edited by Tom. His editorial comments, placing the individual papers in context, along with additional references, make this volume an invaluable reference work. He also edited the Springer *Handbook of Acoustics*, along with serving as author for three of the chapters. This resulted in a truly remarkable compendium. Tom has numerous entries in the *Encyclopedia of Acoustics* published by J. Wiley, in the Macmillan *Encyclopedia of Physics*, in the Focal *Encyclopedia of Electronic Media*, and in the New Grove *Dictionary of Music and Musicians*, where his biography is listed (immediately before Giacomo Rossini). His collaboration with A. J. Houtsma and W. M. Wagenaars in producing a compact disc of Auditory Demonstrations resulted in a marvelous tool for educators. He is the author of over 400 scientific publications, and holds 9 US and 11 foreign patents.

Collaborating with scientists from all corners of the globe, he has hosted young investigators and experienced scientists, thus stimulating research activities in musical acoustics at all levels. Among the many scientists who have passed through his laboratory are such notables as Neville Fletcher, Adrian Houtsma, Johan Sundberg, Bernhard Richardson, and Murray Campbell. In addition to serving as host, he has visited many laboratories and spent time working with colleagues in such places as the Massachusetts Institute of Technology; the Clarendon Lab at Oxford, England; Argonne National Laboratory, Illinois; University of New England, Australia; Royal Institute of Technology, Sweden; Institute for Perception Research, the Netherlands; Physikalisch-Technische Bundesanstalt, Germany; Ecole National Supérieure des Télécommunications, France; Luleå University of Technology, Sweden; University of California, San Diego; Fraunhofer Institute, Germany; University of Edinburgh, Scotland; Stanford University, CA; Seoul National University, Korea; and serving as visiting exchange scholar in China.

Two examples may serve as an indication of Tom's influence in more commercial ventures. Malmark Inc. sought him out to tap his knowledge of bells. This consultation eventually led to the development of an over-sized hand-bell cast in aluminum, the choice of the material dictated by Tom's insight into characteristics of bending wave propagation in bell structures. In another setting, a steel pan artisan in Switzerland, Felix Rohner, of Pan-Art contacted Tom. Insights gained from Tom about vibrational modes eventually led to the development of a new instrument, the "Hang," a hand-played steel instrument, which has gained popularity all over the world.

Tom's influence on Acoustical Society of America (ASA) activities is extensive. As a member of the committees on Education in Acoustics and Musical Acoustics he has organized numerous special sessions at meetings, as well as presenting tutorials on mu-

sical instruments and acoustics demonstrations. His service as editor of *ECHOES* has been truly enlightening, particularly his columns *Scanning the Journals* and *Acoustics in the News*. He also serves as associate editor for education for the *Journal of the Acoustical Society of America* (JASA) and associate editor for musical acoustics for *JASA Express Letters*. He served as chair of the Technical Committee on Musical Acoustics, and as a member of the Medals and Awards Committee and the College of Fellows Steering Committee. His endowment gift to the Acoustical Society Foundation led to the establishment of an ASA prize in education bearing his name.

With all these activities in musical acoustics, Tom did not neglect his deep commitment to quality education. In addition to numerous publications in *The Physics Teacher* and in the *American Journal of Physics*, the two major publications of the American Association of Physics Teachers (AAPT), he published several AAPT resource letters and collections of reprints relevant to physics teaching. He has conducted many teacher workshops, as well as summer courses for physics teachers. He wrote both a laboratory manual to accompany his *Science of Sound* text, and a manual to assist teachers in using the text for a course. He co-authored the text *Light Science* with Chris Chiaverina, where the pun in the title was clearly intentional, as well as a book *Teaching Light and Color*. He has been active in the Minnesota, Illinois, Chicago and California sections of the AAPT, and in 1991 served as national president of the AAPT.

Tom has been the recipient of several awards including Distinguished Research Professor, NIU; National Sigma Xi lecturer; Fellow of the ASA, the American Association for the Advancement of Science, the American Physical Society, and the Institute of Electrical and Electronics Engineers; Honorary membership in the Acoustical Society of India; the AAPT Robert Millikan Medal; the ASA Silver Medal in Musical Acoustics; and Distinguished Service Citations from ASA and AAPT.

Tom is justly proud of his children: Karen, pastor of a Lutheran congregation; Barbara, Professor of New Testament at Lutheran School of Theology at Chicago; Erik, in international business; Jane, Professor of Agricultural and Biological Engineering at Purdue University; and Mary, Mayor of Northfield, Minnesota.

If you plan to visit Tom at his home, you better bring along a musical instrument for an evening playing chamber music, or else be prepared for a great time singing madrigals. Otherwise look for him in the clarinet section of a local orchestra, or listen to him harmonize in a barbershop quartet at an ASA social.

Tom's well deserved international reputation not only makes this award appropriate, but it reflects on the values of a Society which has such deep appreciation for individuals whose contributions are so far reaching.

UWE J. HANSEN
WILLIAM M. HARTMAN

Session 4aAA

Architectural Acoustics and ASA Committee on Standards: Acoustics of Mixed Use Buildings

Steven D. Pettyjohn, Chair
The Acoustics & Vibration Group, 5700 Broadway, Sacramento, CA 95820-1852

Chair's Introduction—8:00

Invited Papers

8:05

4aAA1. Acoustical design of mixed-use buildings: The state of the art (and science). Robert P. Alvarado and Ethan C. Salter (Charles M. Salter Assoc., Inc., 130 Sutter St., Ste. 500, San Francisco, CA 94104, robert.alvarado@cmsalter.com)

Mixed-use developments integrate a variety of uses including residences, offices, shops, restaurants, and theaters into functional, living, and working communities. They are often located near rail, major roadways, and airports, which can generate significant levels of environmental noise and vibration that will need to be mitigated to meet Building Codes and project goals. This paper will detail project-specific issues, solutions, and experiences by Charles M. Salter Associates. Our experience has given us the opportunity to collaborate with various project stakeholders to achieve project goals within the constraints of aesthetics, budget, and space. Authors will present case studies which include acoustical design issues that have been incorporated during the design phase, coordinated during construction, and conveyed to the community that the project serves. Acoustical needs of mixed-use developments can vary based on applicable Codes, as well as project marketing goals and expectations of end users. Educating future owners and tenants about acoustics is important to reducing the design team and the developers' exposure to possible litigation, and also helps to maintain resident health, comfort, and property values. Robert Alvarado has managed hundreds of projects over the past 12 years. His mixed-use project portfolio includes international and domestic projects. Ethan Salter has consulted on dozens of projects throughout the country, including several incorporating sustainable design.

8:25

4aAA2. From an architect's perspective: Acoustical challenges in mixed-use buildings. Dale Farr and Karyn Goodfriend (Fletcher Farr Ayotte, Inc., 520 SW Yamhill, Ste. 900, Portland, OR 97204, kgoodfriend@ffadesign.com)

Incorporating acoustical design in mixed-use buildings can be architecturally challenging. Key issues impacting acoustical design decisions are: (1) the unknown—due to undetermined uses of commercial lease space; (2) the cost—clients do not want to pay for ideal acoustical design because they perceive it to be cost prohibitive. Often clients do not understand the importance of considering acoustic elements during the initial design process until complaints occur much later, postconstruction. Other elements impacting acoustical design are: apartments versus condominiums, existing buildings versus new structures; changes in lease space use; and the possible conflict of the acoustical design not meeting life/safety code. On each architectural project it is ultimately the goal to design to client criteria with a mission to deliver solutions that meet the most needs based on user programs. Often concessions must be made to find the best overall solution for the building and budget. This impacts all of the design consultants, which means designs need to be flexible and are sometimes less than ideal in segregation. It is worth discussing how architects successfully design for the unknown and justify inclusion of acoustical designers early on, while staying within budget.

8:45

4aAA3. Sound and vibration issues encountered in mixed use construction involving residential, office, and retail uses created from a combination of renovation and expansion of an existing space. Steven D. Pettyjohn (The Acoust. & Vib. Group, Inc., 5700 Broadway, Sacramento, CA 95820-1852, spettyjohn@acousticsandvibration.com)

Buildings housing mixed uses often encounter acoustic and vibration problems because expectations differ for the users and patrons. Mixed uses could include retail and office spaces; offices and residential; retail and residential; or, as in the case to be discussed, retail (restaurants) on the first floor, offices on the second floor, and residences on the upper two floors. An existing historic three-story concrete structure originally housing a car dealership in an urban area was renovated to provide all three uses. A fourth floor was added to provide a second floor of residential use. The offices and residential areas were designed to meet interior sound level standards from sources within and exterior to the building. The two restaurants were designed individually by their in house teams with only vibration standards provided by the building design group. The lack of coordination between restaurant and building design teams let to both sound and vibration problems that had to be corrected after construction was mostly complete. Construction teams for the restaurant groups were not familiar with such issues based on their previous installations. This paper discusses the myriad of problems and remedies and provides an outline of the issues that need addressing.

9:05

4aAA4. Low frequency sound problems found in mixed use buildings that house entertainment venues and residential developments and containment options. Scott W. Smith (Ballentine Walker Smith, Inc., Kennesaw, GA 30144, bwsacoustics@bellsouth.net) and Steven D. Pettyjohn (The Acoust. & Vib. Group, Inc., Sacramento, CA 95820-1852)

Many problems are to be expected when mixed use buildings include restaurants and residential spaces. When the residential spaces are condominiums and the restaurant becomes a nightclub, the sound problems multiply quickly. The low frequency sound produced in a nightclub featuring music catering to a young crowd is of particular concern. This is partially because of the difficulty of finding remedies once the building construction is complete. This is the situation that arose in a facility recently completed. The nightclub wanted to continue its operation while the condominium owners wanted a resolution of the problem. Sound tests were completed in the residential spaces during operation of the nightclub, but the low frequency content was not always the same, requiring multiple attempts to measure in the source and receiving spaces. Results of these measurements and the recommendations for correcting the problem are presented in this paper. The goal is to provide results of sound measurements made after the recommendations are implemented. Again, sound will be measured in the source and receiver spaces to understand how the noise reduction changed and compared with the predicted sound reduction.

Contributed Paper

9:25

4aAA5. An historic conversion: From a bank to a restaurant and residences. Ioana Pieleanu, Jeffrey Fullerton, and Benjamin Markham (Acentech Inc., 33 Moulton St., Cambridge, MA 02135)

A conversion of an old bank building in Boston's tony South End to a mixed-use building featuring retail on the ground floor and luxury condominiums above was completed in 2007. More recently, a new restaurant (garnering awards for its interior design and rave reviews for its food) has opened in one of the ground floor retail spaces directly below a particularly noise-sensitive resident. Consultants at Acentech worked on two aspects of the project: first, on the base building as consultants to the architect, and second, on the isolation between the restaurant and the second floor residences as consultants to the restaurant. Using this case study and extensive data measured on site, the authors will discuss best practices to achieve good sound isolation in mixed-use buildings, common pitfalls that result from working with existing historic structures, and some difficulties in achieving the high degree of sound isolation that some luxury condominium owners expect.

THURSDAY MORNING, 21 MAY 2009 GALLERIA NORTH, 8:25 TO 11:50 A.M.

Session 4aAB

Animal Bioacoustics: General Topics in Animal Bioacoustics I

Holger Klinck, Chair
CIMRS, Oregon State Univ., Newport, OR 97365

Chair's Introduction—8:25

Contributed Papers

8:30

4aAB1. Auditory temporal summation in pinnipeds. Asila Ghoul (Univ. of California Santa Cruz Long Marine Lab., 100 Shaffer Rd., Santa Cruz, CA 95060), Marla M. Holt (Natl. Marine Fisheries Service, Seattle, WA 98112), Colleen Reichmuth, and David Kastak (Univ. of California Santa Cruz Long Marine Lab., Santa Cruz, CA 95060)

In addition to improving the understanding of auditory processing in pinnipeds, direct measures of temporal summation are relevant to the selection of signal parameters when conducting audiometric research, assessing the effects of signal duration on communication ranges, and evaluating the potential auditory impacts of anthropogenic signals. In the present study, individuals from three pinniped species were tested to determine how signal duration influenced pure-tone hearing thresholds. The psychophysical method of constant stimuli was used to obtain aerial thresholds for each subject at nine different signal durations ranging from 25 to 500 ms. Parameter estimates derived for a California sea lion (*Zalophus californianus*) from an exponential model of temporal summation yielded time constants (τ) of 176, 98, and 141 ms at frequencies of 2.5, 5, and 10 kHz, respectively. Preliminary results with a northern elephant seal (*Mirounga angustirostris*) at 5 kHz (this study), and a harbor seal (*Phoca vitulina*) at 2.5 kHz [M. M. Holt et al., J. Soc. Am. **116**, 2531 (2004)] show similar values for (τ), 134 and 144 ms, respectively. These time constants are similar to those of other mammals tested and do not appear to vary with respect to frequency.

8:45

4aAB2. Annual temporal patterning in the vocalizations of captive seals: Two long-term case studies. Colleen Reichmuth and Ronald J. Schusterman (Inst. of Marine Sci., Univ. of California Santa Cruz, 100 Shaffer Rd., Santa Cruz, CA 95060)

Seasonal changes in vocalizations occur in a variety of species. Factors such as the condition of conspecifics, physiological states that in turn may be related to environmental cues, and developmental and individual differences all potentially influence temporal changes in sound production. In the present study, the vocal behavior of two captive seals was monitored daily for over 10 yrs. Both seals were housed in the absence of conspecifics from the age of 1 yr extending past sexual maturity. The male harbor seal (*Phoca vitulina*) began characteristic underwater vocal displays at the age of 6. Intense periods of acoustic activity lasted weeks to months, overlapped with the breeding activity of local harbor seals, and comprised stereotypic sound emissions that were structurally similar to those reported for wild seals. The female northern elephant seal (*Mirounga angustirostris*) produced aberrant intense airborne vocalizations from the age of 4 that were annually synchronized to a period of approximately 5 weeks coinciding with estrous. Endogenous changes appear to trigger these behavioral cycles, presumably as a

result of hormonal changes associated with photoperiod. Vocalizations may be a noninvasive indicator of reproductive state and therefore may provide a useful management and conservation tool in captive settings.

9:00

4aAB3. A comparison of behavioral and electrophysiological measures of aerial hearing sensitivity in a Steller sea lion (*Eumetopias jubatus*). Jason Mulsow (Dept. of Ocean Sci., Univ. of California Santa Cruz, Earth and Marine Sci. Bldg., Santa Cruz, CA 95064) and Colleen Reichmuth (Univ. of California Santa Cruz, Santa Cruz, CA 95060)

A number of studies with odontocete cetaceans have demonstrated that hearing sensitivity measurements using electrophysiological auditory steady-state responses (ASSRs) can provide an efficient means of estimating a subject's behavioral audiogram. Expansion of ASSR methods to another marine mammal group, the otariid pinnipeds (sea lions and fur seals), holds the potential to increase the number of otariid individuals and species for which hearing sensitivity data are available. A within-subject comparison of ASSR and behavioral measures of aerial hearing sensitivity was conducted with an individual of the largest otariid species, the Steller sea lion. Psychophysical methods were used to obtain an unmasked aerial audiogram at 13 frequencies spanning a range of 0.125 to 34 kHz. Corresponding ASSR thresholds measured at frequencies of 1, 2, 5, 10, 20, and 32 kHz had differences (relative to behavioral thresholds) ranging from 1 dB at 20 kHz to 30 dB at 1 kHz. Overall, the ASSR audiogram was a fairly accurate predictor of the behavioral audiogram at frequencies of 2 kHz and above. Our results suggest that ASSR methods can be appropriately applied to otariid pinnipeds in estimating aerial sensitivity at frequencies of approximately 2 kHz and above.

9:15

4aAB4. Vibration characteristics of the tympanoperiotic complex in the bottlenose dolphin, *Tursiops truncatus*. Petr Krysl (Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0085), Ted W. Cranford (San Diego State Univ., San Diego, CA 92182), and John A. Hildebrand (Univ. of California, San Diego, La Jolla, CA 92093-0205)

Modal finite + boundary element analysis of a bottlenose dolphin's bony tympanoperiotic complex, including the ossicles, was performed to determine the mode shapes and natural frequencies. The goal was to gain insight into the transmission of sound pressure waves arriving through the soft tissues and transmitted across the bony components into the oval window of the inner ear. The finite element model of the bones was derived from CT scans with a 360 μm voxel resolution. In the first approximation the soft tissue was considered to be acoustically equivalent to an incompressible inviscid liquid, taken as infinite in extent. The added mass terms were computed with a boundary element model. The computed frequencies cover the range up to 160 kHz. The capacity of the natural vibration modes to excite motion of the stapes footplate was assessed by measuring the relative motion of the incudostapedial joint normalized by the normal displacement of the wet-surface of the ear bones. In addition to a quantitative assessment a number of qualitative observations may be made that could explain the function of the dolphin's ear complex. For example, the vibrational patterns are nontrivial and frequency dependent. [Work supported by the U.S. Navy CNO45.]

9:30

4aAB5. "Rivers" of sound in Cuvier's beaked whale (*Ziphius cavirostris*): Implications for the evolution of sound reception in odontocetes. Ted W. Cranford (Biology Dept., San Diego State Univ., 2674 Russmar, San Diego, CA 92182), Petr Krysl, and John A. Hildebrand (Univ. of California at San Diego, La Jolla, CA 92093)

Industrial CT scanning technology was used to collect the first x-ray tomograms from the head of an adult male Cuvier's beaked whale. These scans and tissue property measurements were used to construct a finite element model. Simulations revealed pathways for sound propagation into and out of the head. One intriguing result concerns a newly described gular pathway by which sound reaches the hearing apparatus. Propagated sound waves enter the ventral aspect of the head and form an acoustic "river" that flows toward the bony ear complexes through the internal mandibular fat bodies. The precise pathway and dimensions of the sound river vary with frequency, but it converges on the bony tympanoperiotic complex. A combination of tissue structures and air spaces act like an internal acoustic pinna that filters and concentrates the incoming sound. The river of sound apparently functions in concert with the absence of the medial bony lamina of the posterior portion of the mandible, a condition that exists in all toothed whales and their ancestral archaeocetes. The gular pathway and river of sound suggests that this is the primordial pathway for underwater hearing in whales and that Norris' jaw hearing mechanism was a more recent development.

9:45

4aAB6. Dall's porpoise (*Phocoenoides dalli*) echolocation click spectral structure. Hannah R. Bassett, Simone Baumann, Gregory S. Campbell, Sean M. Wiggins, and John A. Hildebrand (Marine Physical Lab, Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093, hbassett@ucsd.edu)

Dall's porpoise (*Phocoenoides dalli*) echolocation clicks have not been widely recorded. Concurrent with visual observations, acoustic recordings of free-ranging Dall's porpoise were made offshore of southern California using a towed hydrophone array with two elements of 250 kHz bandwidth. We examined 6035 clicks from 12 sessions totaling more than two hours over the course of seven days. The Dall's porpoise echolocations recorded were short (48–804 μs), narrow band (2–10 kHz [−3dB]) clicks with most peak frequencies between 117 and 141 kHz, but some as high as 198 kHz. Many clicks contained a multipulse temporal structure, resulting in stereotyped spectral peaks and notches. Two distinctive click types with different spectral banding patterns and peak frequencies (122.8 and 135.8 kHz) were observed. Spectral banding patterns have been used as a species identifier for Risso's dolphins and Pacific white-sided dolphins. These two dolphins and Dall's porpoise have similar head morphologies, which may play a role in producing clicks with spectral peaks and notches. This study shows that Dall's porpoise produce multiple click types, which may provide a tool for population classification, and that their clicks contain spectral banding patterns, which may provide insight into the mechanism by which such clicks are produced.

10:00—10:20 Break

10:20

4aAB7. Analysis of most prominent signal features of humpback whale (*Megaptera Novaeangliae*) vocalizations towards the goal of autonomous acoustic classification. Ted Abbot, Owen Mayer, Vince Premus, Philip Abbot, and Ira Dyer (OASIS, Inc., 5 Militia Dr. Lexington, MA 02421)

Humpback whale vocalizations were recorded using hydrophones on glider systems off Alaska in January 2000, in Hawaii in February 2008, and in the Stellwagen Bank National Marine Sanctuary in October 2007 and July 2008. The vocalizations have been grouped into five call types based on the most prominent signal features. Only five call types are used because autonomous species classification relies on the most consistent and repeatable signal features rather than the full diverse range of humpback vocalizations. The five call types are upsweep (increasing frequency over time), downsweep (decreasing frequency over time), flute (increasing and decreasing frequency over time), tone (little or no change in frequency over time), and groan (commonly a social or feeding-related vocalization, frequently characterized by unstructured broadband sound). We present detailed statistical analyses of these call types including bandwidth, minimum and maximum frequency, duration, and slope. A comparative analysis across data sets shows the relative frequency of occurrence of each vocalization type and indicates the degree of temporal and geographic variation of Humpback vocalizations.

10:35

4aAB8. Is rejection of clutter achieved by disrupting perception of delay in bat sonar? Mary E. Bates (Dept. of Psych., Brown Univ., 89 Waterman St., Providence, RI 02912) and James A. Simmons (Brown Univ., Providence, RI 02912)

Big brown bats (*Eptesicus fuscus*) emit frequency-modulated (FM) biosonar sounds containing two prominent harmonics, FM1 (55–22 kHz) and FM2 (105–145 kHz), and perceive target distance from echo delay. Ordinarily, echoes from objects arrive with both harmonics at the same delay, although FM2 becomes progressively more attenuated than FM1 with increasing target distance and off-axis direction. Off-axis or long-range echoes naturally undergo lowpass filtering, which causes significant loss of acuity. Misalignment of FM2 with respect to FM1 or selective removal of low-end frequencies disproportionately affects the accuracy of delay perception. Delay acuity is sharp for echoes containing both harmonics, less sharp for echoes containing only FM1, and very poor for echoes containing only FM2. Acoustically unnatural highpass filtering to remove low-end frequencies from FM1 causes acuity to collapse. Attenuation of FM2 relative to FM1 by 3 dB decreases delay acuity, but shortening of the delay for FM2 by 48 μs counteracts amplitude-latency trading and restores delay acuity. Bats may have a spatial perceptual "fovea" covering a narrow zone in front of them for high-acuity perception surrounded by a zone of much lower acuity that suppresses the perceptual salience of background clutter. [Work supported by NASA RI Space Grant and ONR.]

10:50

4aAB9. Range discrimination of multiple objects from the echo spectrogram measured by using the frequency modulation sound. Ikuo Matsuo (Dept. of Information Sci. Tohoku Gakuin Univ., Sendai, 981-3193, Japan, matsuo@cs.tohoku-gakuin.ac.jp)

Using the echolocation, bats can capture moving objects in real 3D 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 to determine delay times of multiple objects from the simulated echoes by using the frequency modulation sound have been previously proposed. In order to extract the temporal changes, Gaussian chirplets with a carrier frequency compatible with emission sweep rates were used. The delay time for first object can be estimated from the echo spectrum around the onset time. The delay time for second object is obtained by adding the delay time for the first object to the delay separation between first and second objects (extracted from the first appearance of interference effects). Further objects can be located in sequence by this same procedure. In this paper, these echoes were measured from multiple closely spaced objects, and it was examined that this model could estimate each delay times of these objects.

11:05

4aAB10. The acoustic implications of sella length in horseshoe bats. Rolf Mülller (Dept. of Mech. Eng., Virginia Tech. & Inst. for Adv. Learning and Res., 150 Slayton Ave., Danville, VA 24540, rolf.mueller@vt.edu), Zhiwei Zhang (Shandong Univ., Hongjia Lou 5, 250100 Jinan, China), and Son Nguyen Truong (Vietnamese Acad. of Sci. and Technol., Hanoi, Vietnam)

Horseshoe bats emit their biosonar pulses through nostrils that are surrounded by elaborate baffle shapes (noseleaves). The shape and size of the different noseleaf shape elements can vary considerably between species. Here, the acoustic effects of the length of the sella have been investigated based on the noseleaf of Bourret's Horseshoe Bat (*Rhinolophus paradoxolophus*) which features an exceptionally long sella: High-resolution digital representations of the noseleaves of three specimens were used to vary the sella length below and above its natural value. A principal axis through the sella was used to define the direction of this scaling. The acoustic beamforming properties of the noseleaf were then assessed as a function of sella scale using numerical methods. It was found that elongating the sella narrows the biosonar beam of the second harmonic in elevation. Furthermore, the natural length of the sella in Bourret's Horseshoe Bat coincided with a fiducial point in the relationship between sella length and beamwidth, where the effect of sella length on beamwidth saturates. This allows the prediction of the natural sella length from its acoustic properties. The long sella in this species could, hence, be an adaptation to producing a narrow beam at comparatively low frequencies.

11:20

4aAB11. An evolutionary approach to perissodactyl vocalizations. David G. Browning (Phys. Dept., Univ. of Rhode Island, 2 Lippitt Rd., Kingston, RI 02881, decibeldb@aol.com) and Peter M. Scheifele (Univ. of Cincinnati, Cincinnati, OH 45267-0379)

Approximately 65 million years ago, dinosaurs were displaced by mammals. In a then highly forested world, the perissodactyls (odd-toed ungulates) became the largest group of large mammals, consisting of an estimated 15 families. We hypothesize that they started to develop a melodic component in their tonal vocalizations in order to provide identity in this restricted visibility setting. With time, large areas shifted to grassland, favoring mammals with compound stomachs, namely, artiodactyles (sheep, cattle, etc.) and resulting in the decline of the perissodactyls to only three families (equines, tapirs, and rhinos) today. Although all the remaining perissodactyls still retain a melodic component in their vocalizations, it appears to be less developed in those that adapted to the open grassland environment, such as the plains zebra. There a simple tonal call (as the artiodactyles have, too) is sufficient to draw attention; additional information can be obtained visually. On the other hand, those that continued in a forested location such as the Sumatran Rhino have developed a more lyrical call to compensate for reduced visibility.

11:35

4aAB12. Time domain, frequency domain, and spectrogram analysis of baleen whale songs. Pranab K. Dhar and Jong-Myon Kim (Univ. of Ulsan, 680-749, Korea, jongmyon.kim@gmail.com)

Whale song is the sound to communicate, and it shows a specific pattern of regular and predictable sounds made by some species of whales. In this study, different baleen whale songs were analyzed in terms of time domain, frequency domain, and spectrogram representation. More specifically, each whale song in the species was analyzed with a sound analysis tool in terms of peak frequency (frequency with peak energy), frequency range of song, and pattern of song production. Cordell North Canyon humpback whales produce the highest peak frequency (1,348 Hz) whereas North Eastern Pacific blue whales produce the lowest peak frequency (18 Hz). The humpback whales provide wide frequency range from 600 Hz to 2.8 kHz, but Atlantic Ocean fin whales provide short frequency range from 15 to 40 Hz. Patterns of song production for the species were also analyzed. Minke and right whales generate similar repeated song. However, humpback, bowhead, blue, and fin whales generate same repeated song. These evaluation techniques can provide solutions for characterizing specific features of whale songs. [Work supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Assessment) (IITA-2008-(C1090-0801-0039)).]

Session 4aBB

Biomedical Ultrasound/Bioresponse to Vibration: Image Enhancement and Targeted Drug and Gene Delivery

Azzdine Y. Ammi, Cochair
Oregon Health and Science Univ., Portland, OR 97239

Saurabh Datta, Cochair
Dept. of Biomedical Engineering, Univ. of Cincinnati, Cincinnati, OH 45242-0586

Invited Papers

8:00

4aBB1. Sonothrombolysis for acute coronary syndromes: Opportunities and challenges. Sanjiv Kaul (Cardiovascular Div., Oregon Health & Sci. Univ., Portland, OR 97239)

Percutanous interventions and pharmacological thrombolysis are the current options for treatment of acute coronary syndromes (ACS). The former is limited by its availability and the latter by its efficacy. Sonothrombolysis has been demonstrated to be effective in achieving tissue perfusion in the peripheral arteries as well as in cerebral arteries. Therefore, there is potential of using sonothrombolysis for the treatment of ACS. An ultrasound imaging and delivery system could overcome the issue of access while combining it with microbubbles and low dose thrombolytics could result in a high reperfusion rate. Furthermore, the direct effect of ultrasound on ishemic myocardium (release of nitric oxide and increase in myocardial blood flow despite total coronary occlusion) could be exploited to protect the myocardium until reperfusion has been achieved. The success of reperfusion could be assessed in real-time using microbubbles. In order to achieve these goals, we need to plan systematic *in vitro* and *in vivo* studies to better understand the mechanics of sonothrombolysis. We also need to develop 4D combined imaging and ultrasound delivery systems. Positive developments in this field can translate into a major impact on human health world wide.

8:20

4aBB2. Targeted microbubble technology and ultrasound-mediated gene delivery. Jonathan Lindner (Cardiovascular Div., Oregon Health and Sci. Univ., 3181 S.W. Sam Jackson Park Rd., Portland, OR 97239, lindnerj@ohsu.edu)

There is interest in harnessing the energy from ultrasound-microbubble interactions for therapeutic gain. Microbubbles (MB) can be used as vectors for ultrasound mediated gene delivery (UMGD). This process relies on coupling cDNA to MB that undergo cavitation resulting in delivery/transfection from controlled bioeffects. Proximity of gene-laden MB to the vessel wall is a critical determinant of UMGD, which may be enhanced by the use of targeted MB probes that bind to certain disease states. In this presentation, endothelial targeting strategies will be discussed together with their recent application to optimize UMGD. Flow chamber and intravital microscopy studies have been performed that demonstrate that the coupling of cDNA and MB surfaces does not interfere with binding of these agents to counterligands on the vascular endothelium. Imaging experiments have also been performed demonstrating that MB retention in tissue using an ischemia-targeting approach is possible with gene-laden MBs targeted to endothelial cell adhesion molecules. Finally, data on the relative transfection efficiency of reporter genes with UMGD using a targeted versus control MB vector approach will be discussed. The overall conclusion of the presentation is that the ability to target MB to disease-related molecules may increase both efficiency and specificity of tissue transfection.

8:40

4aBB3. Gas body destruction reduces the effective circulating dose of ultrasound contrast agent infused in rats. Douglas L. Miller, Chunyan Dou (Dept. of Radiology, Univ. of Michigan, 1301 Catherine St., Ann Arbor, MI 48109-5667, douglm@umich.edu), and Roger C. Wiggins (Univ. of Michigan, Ann Arbor, MI 48109)

Glomerular capillary hemorrhage (GCH) in rat kidney provides a model system for assessing *in vivo* gas body efficacy in diagnostic or therapeutic applications of ultrasound. Two diagnostic ultrasound machines were utilized: one monitored the second-harmonic B mode contrast-enhancement of the left kidney and the other exposed the right kidney for GCH production. Definity contrast agent was infused at 5 μl/kg/min for 300 s during shams and 1.5-MHz intermittent exposures at 2.3-MPa peak rarefactional pressure amplitude in groups of five rats. The left kidney image brightness enhancement, indicative of circulating gas body dose, was 18.4 au (decompressed arbitrary units) in shams with no GCH in histology. Exposure of the right kidney with a normal 1-s image interval induced 68.4% GCH but reduced the left kidney enhancement to 3.3 au, which implies substantial gas body destruction. Decreased exposure with 10-s interval reduced right kidney GCH ($P' < 0.001$) but only to 30.3% while ameliorating gas body destruction with 13.1-au left kidney enhancement. The effective *in vivo* gas body dose in rats may be reduced greatly due to gas body destruction in the small animal, complicating extrapolation to similar conditions of human exposure. [Work supported by NIH grant EB00338.]

9:00

4aBB4. Combined effect of ultrasound and liposomal doxorubicin on AT2 Dunning tumor growth in rats: Preliminary results. Lucie Somaglino (Inserm U556, 151 cours Albert Thomas, 69424 Lyon cedex 03, France and Univ. de Lyon, Lyon F-69003, France, lucie.somaglino@inserm.fr), Guillaume Bouchoux, Sabrina Chesnais, Anis Amdouni, Jean-Louis Mestas (Inserm U556, 69424 Lyon cedex 03, France), Sigrid Fossheim, Esben A. Nilssen (Epitarget AS, 0307 Oslo, Norway), Jean-Yves Chapelon, and Cyril Lafon (Inserm U556, 69424 Lyon cedex 03, France)

Previous in vitro studies conducted in our group have shown the feasibility of monitoring drug release from liposomes by an inertial acoustic cavitation index. We currently report in vivo experiments utilizing the cavitation index in combined treatment of AT2 Dunning tumor grafted rats with focused ultrasound and liposomal doxorubicin. Sixty-three rats were allocated into seven groups: control, low level ultrasound treatment, high level ultrasound treatment, free doxorubicin+high level ultrasound treatment, and liposomal doxorubicin, liposomal doxorubicin+low level ultrasound treatment, and liposomal doxorubicin+high level ultrasound treatment. Based on pharmacokinetic studies, it was decided to apply ultrasound to the tumor 48 h after drug injection. An experimental setup was built to perform repeatable and rapid sonications of tumors monitored by the cavitation index. Tumor growth was assessed for a period of 35 days after tumor inoculation. Results showed that liposomal doxorubicin significantly slowed down tumor growth. However, the synergy between ultrasound and liposomal doxorubicin could not be firmly demonstrated. The lack of synergy may be due to inefficient induction of drug delivery in vivo or too high liposome dosage hiding synergistic effects. [Work funded by the Norwegian Research Council (NANOMAT programme). Epitarget AS is acknowledged for the supply of liposomes.]

9:15

4aBB5. Ultrasound-enhanced drug delivery through sclera. Robin Shah and Vesna Zderic (Dept. of Elec. and Comp. Eng., The George Washington Univ., 801 22nd St. NW, Washington, DC 20052, zderic@gwu.edu)

Achieving an increase in drug delivery through the sclera is important in the treatment of the back of the eye diseases including macular degeneration, diabetic retinopathy, etc. Our objective is to utilize therapeutic ultrasound in enhancing drug delivery through the sclera. Porcine sclera was placed in a standard diffusion cell at a normal physiological temperature of 34 °C. Solution of sodium fluorescein, a hydrophilic drug-mimicking compound, was added to donor compartment, and receiver compartment was filled with saline. The sclera was exposed to ultrasound for 5 min (intensities 1.2–1.8 W/cm^2 and frequencies 0.5 to 5 MHz). After 60 min, solution samples were taken from the receiver compartment to determine the concentration of sodium fluorescein. The sclera permeability to the drug mimicking agent *in vitro* increased 3.5 times at 0.5 MHz (p-value of less than 0.05), 1.7 times at 1 MHz, 3.5 times at 3.5 MHz (p-value of less than 0.05), and 1.5 times at 5 MHz. The average temperature of the sclera during ultrasound exposure was 42 °C. No gross changes were observed in the sclera due to ultrasound application. Future work will focus on determination of optimal ultrasound parameters for the drug delivery through the sclera.

9:30

4aBB6. Direct observation of microbubble interactions with ex vivo microvessels. Hong Chen, Andrew A. Brayman, Michael R. Bailey, and Thomas J. Matula (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105, hongchen@apl.washington.edu)

The interaction between microbubbles with tissue is poorly understood. Experimental evidence, supported by numerical simulations, suggests that bubble dynamics is highly constrained within blood vessels. To investigate this further, a high-speed microimaging system was set up to study the effects of acoustically activated microbubbles on microvessels in ex vivo rat mesentery tissues. The microbubble-perfused tissues were placed under a microscope and insonified with MHz ultrasound. A variety of interactions was observed by a high-speed camera: arterioles, venules, and capillaries were all recorded to dilate and invaginate by activated microbubbles. For small diameter microvessels, dilation and invagination were nearly symmetric, and bubble-induced rupture of the vessel was observed at high pressure. For larger microvessels, the portion of the vessel nearest the bubble coupled the strongest to the bubble dynamics, and the extent of dilation was smaller than invagination. Tissue jetting toward the bubble was recorded in many cases. The interaction of multiple bubbles inside microvessels was also observed. Bubble oscillation, vessel wall velocity, and tissue jet velocity were quantitatively measured. Invagination of vessel walls, especially tissue jetting, may be the major mechanism for tissue injury by a bubble. [Work supported by NIH 5R01EB000350.]

9:45

4aBB7. Shell buckling enhances subharmonic behavior of phospholipid coated ultrasound contrast agent microbubbles. Jeroen Sijl, Timo Rozendal, Marlies Overvelde, Valeria Garbin, Benjamin Dollet, Nico de Jong, Detlef Lohse, and Michel Versluis (Phys. of Fluids Group, Univ. of Twente, Enschede, The Netherlands)

Subharmonic behavior of coated microbubbles can greatly enhance the contrast in ultrasound imaging. The threshold driving pressure above which subharmonic oscillations are initiated can be calculated from a linearized Rayleigh-Plesset-type equation. Earlier experimental studies on a suspension of phospholipid-coated microbubbles showed a lower threshold than predicted from traditional elastic shell models. Here we present an experimental study of the subharmonic behavior of individual BR-14 microbubbles (Bracco Research) with initial radii between 1.6 and 4.8 μm. The subharmonic behavior was studied as a function of the amplitude and the frequency of the driving pressure pulse. The radial response of the microbubbles was recorded with the Brandaris ultrahigh-speed camera, while the resulting acoustic response was measured with a calibrated transducer. It is shown that the threshold pressure is minimum near a driving frequency equal to half the resonance frequency of the bubble, as expected. We found a threshold pressure as low as 10 kPa for certain bubble sizes, which can be explained by the shell buckling model proposed by [Marmottant *et al.*, JASA (2005)]. We show that the origin of subharmonic behavior is a result of the discontinuous transition within the bubble shell from the elastic state to the tensionless buckling state.

10:00—10:30 Break

10:30

4aBB8. Shell buckling increases the nonlinear dynamics of ultrasound contrast agents at low acoustic pressures. Marlies Overvelde (Phys. of Fluids, Univ. of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands, m.l.j.overvelde@utwente.nl), Benjamin Dollet, Valeria Garbin (Univ. of Twente, Enschede, The Netherlands), Nico de Jong (Experimental Echocardiography, Thoraxcenter, Erasmus MC, Rotterdam, The Netherlands), Detlef Lohse, and Michel Versluis (Univ. of Twente, Enschede, The Netherlands)

The key feature of ultrasound contrast agents in distinguishing blood pool and tissue echoes is based on the nonlinear behavior of the bubbles. Here we investigate the nonlinear properties of the shell which lead to an increased nonlinear bubble response, especially at low acoustic pressures. The microbubbles were studied in free space away from the wall using the Brandaris camera coupled to an optical tweezers setup. The microbubble spectroscopy method [Van der Meer et al., JASA, **121**, 648 (2007)] was employed to characterize BR-14 microbubbles (Bracco, Geneva). For increasing applied pressures the bubble resonance curves become asymmetrical and the frequency of maximum response decreases, up to 50% at a pressure of 25 kPa. It was found that the skewing of the nonlinear resonance curve is the origin of the so-called thresholding behavior below resonance. Traditional bubble models account for a purely elastic shell predict linear behavior, whereas the shell buckling model by Marmottant et al. [JASA, **118**, 3499

(2005)], originally developed to predict compression-only behavior, captures the asymmetry of the resonance curve in great detail. The full understanding of the nonlinear behavior at low acoustic pressures opens a wealth of new possibilities in contrast-enhanced ultrasound imaging.

10:45

4aBB9. Influence of finite wall impedance on contrast agent bubble behavior near a membrane. Todd Hay, Marlies Overvelde, Benjamin Dollet, Valeria Garbin, Nico de Jong, Detlef Lohse, and Michel Versluis (Phys. of Fluids, Univ. of Twente, 7500 AE Enschede, The Netherlands)

Experiments investigating the radial dynamics of ultrasound contrast agent (UCA) microbubbles in the vicinity of an optically transparent membrane show that bubble oscillation amplitude and frequency of peak response decrease as bubbles move closer to the membrane (Overvelde *et al.* Proc. 19th Intl. Congr. Acoust., 2007). However, treating the membrane as a rigid wall predicts that the peak oscillation amplitude should increase at small standoff distances, contrary to experimental observations. Here we present a model describing UCA bubble dynamics near a locally reacting wall having finite acoustic impedance. If finite wall impedance is included in the model, the predicted bubble behavior is in good agreement with observations. The hybrid time-frequency domain model is based on a linear frequency domain solution [Ingard, J. Acoust. Soc. Am. **23**, 329 (1951)] which has been adapted to account for weakly nonlinear bubble oscillations and UCA shell dynamics. Wall impedance parameters are derived from independent experimental measurements. Comparisons between the model and data from microbubble spectroscopy experiments will be presented. [Work supported by the ASA Hunt Fellowship and NIH DK070618.]

11:00

4aBB10. Model for bubble dynamics in liquid near an elastic tissue interface. Todd A. Hay (Phys. of Fluids, Univ. of Twente, 7500 AE Enschede, The Netherlands), Evgenia A. Zabolotskaya, Yurii A. Ilinskii, and Mark F. Hamilton (The Univ. of Texas at Austin, Austin, TX 78713-8029)

A model is under development for the weakly nonlinear dynamics of a bubble in a blood vessel surrounded by tissue with elasticity and losses. The model requires knowledge of the radiation impedance of the bubble within the vessel, which is determined by the Green's function. As a first step toward this objective, the Green's function is derived for a bubble in a liquid half-space bounded by an elastic half-space. The method used to derive the Green's function follows an approach described previously for the response of an acoustically driven object in an elastic half-space [Zabolotskaya *et al.*, J. Acoust. Soc. Am. **124**, 2514(A) (2008)]. The Green's function is expressed in terms of its angular spectrum in a plane parallel to the interface, resulting in ordinary differential equations in the coordinate normal to the interface. Boundary conditions are satisfied at the interface and in the plane containing the source to obtain solutions of the differential equations, the inverse spatial Fourier transform of which yields the desired Green's function. Extension to multiple layers of tissue is straightforward. Calculations will be presented for the radiation impedance of the bubble near the interface. [Work supported by the ASA Hunt Fellowship and NIH DK070618.]

11:15

4aBB11. Determination of maximum response radius at therapeutic pressure levels. Kelsey J. Carvell and Timothy A. Bigelow (Dept. of Elec. and Comput. Eng., Iowa State Univ., 2011 Coover Hall, Ames, IA 50010, kcarvell@iastate.edu)

High-intensity, focused ultrasound therapy is a minimally invasive therapy technique that has shown potential in many therapeutic applications, especially when coupled with the cavitation of microbubbles. The purpose of this study was to determine the effect of pressure amplitude on cavitation resonance frequency/bubble size at therapeutic field levels. Earlier work has indicated that the resonance frequency depends on pressure amplitude; however, the investigation only considered pressure amplitudes up to 1 MPa [Ultrasonics **43**, 113 (2004)]. Our study was conducted by simulating the response of bubbles using the Gilmore-Akulichev formulation to solve for the bubble response. The frequency of the sine wave varied from 1–5 MHz while the amplitude of the sine wave varied from 0.0001–13 MPa. The resonance size for a particular frequency of excitation and amplitude was determined by finding the initial bubble size that resulted in the maximum bubble expansion relative to the initial size for an air bubble in water prior to the first collapse. Preliminary simulations indicated that this metric gave the correct resonance size for small excitations. These simulations demonstrated a downshift in resonance size with increasing pressure amplitude.

11:30

4aBB12. Subharmonic response from single, polymer-shell contrast agents subjected to high-frequency excitation. Parag V. Chitnis, Jonathan Mamou, Paul Lee, and Jeffrey A. Ketterling (Lizzi Ctr. for Biomedical Eng., Riverside Res. Inst., 156 William St., 9th Fl., New York, NY 10038)

Three polymer-shelled agents (nominal mean diameters of 0.56, 1.1, and 3.4 μm) from POINT Biomedical were investigated to determine the optimal parameter space for generating a subharmonic response from single agents when subjected to high-frequency (above 20 MHz) excitation. A flow phantom was constructed to restrict the flow of a dilute contrast agent solution to a small volume. Two single-element transducers, with nominal center frequencies of 40 and 20 MHz, were aligned such that they were confocal and orthogonal to each other, and their mutual focus was positioned within the flow phantom. The 40 MHz transducer was used in transmit/receive mode while the 20 MHz transducer was used passively in the receive mode. The radio-frequency backscatter signals from individual contrast agents were digitized simultaneously from both transducers under a variety of pulse durations (5–20 cycles), acoustic pressure levels (1.5–6.0 MPa), and driving frequencies (35–45 MHz). Echo signals from individual contrast agents were windowed and spectra were calculated. For each contrast agent, the magnitude of the subharmonic (20–MHz) response was normalized with respect to the magnitude of the fundamental (40–MHz) backscatter. Experimental results agreed closely with theoretical calculations. [Work supported by NIH EB006372.]

11:45

4aBB13. Dissolution and stabilization of contrast microbubble by its elastic shell. Kausik Sarkar and Amit Katiyar (Mech. Eng., Univ. of Delaware, Newark, DE 19716)

We have developed a model for gas transport from encapsulated contrast microbubbles used for ultrasound imaging and drug delivery. The model explicitly accounts for permeability of gas through encapsulation and encapsulation elasticity. We use it to investigate dissolution time and stability of a lipid-coated perfluorocarbon bubble such as Definity. Realistic values of material properties such as diffusivities, permeabilities, and Ostwald coefficients of air and octafluoropropane were used to find that the lifetime of hours for such a bubble is only possible at extremely low surface tension in the absence of shell elasticity. Dissolution dynamics is investigated for effects of various parameters such as initial mole fraction of octafluoropropane, initial radius, surface tension, and shell permeability. The dissolution dynamics scales with permeability, in that when the time is nondimensioanlized with permeability, curves for different permeabilities collapse on a single curve. For an elastic shell, one obtains stable bubbles when the appropriate condition on surface tension and encapsulation elasticity is satisfied. The condition varies with the level of air saturation of the surrounding medium. For the oversaturated case, the dynamics allows multiple branches of equilibrium solutions. Stability of these branches are numerically investigated and discussed.

Session 4aID

Interdisciplinary: Workshop on Preparing Articles for the Journal of the Acoustical Society of America (JASA) and JASA Express Letters

Ning Xiang, Chair
Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180

Chair's Introduction—8:55

Invited Papers

9:00

4aID1. Achieving publication excellence in the Journal of the Acoustical Society of America. Ning Xiang (Graduate Prog. in Architect. Acoust., Rensselaer Polytech. Inst., Troy, NY 12180)

Publications in refereed acoustics journals are of significant relevance for scientists and acousticians in the acoustics-related fields. The Journal of the Acoustical Society of America (JASA) encourages authors to submit papers for JASA publication. In achieving publishing excellence, the JASA regularly publishes detailed, updated guidance and instructions. As a frequent reviewer, an author of JASA papers, and an Associate Editor of the JASA, this paper will provide an outline and overview of its peer-review process of the JASA to analyze possible reasons for successful and unsuccessful publication effort. This paper also discusses how to prepare qualified manuscripts, how to avoid unnecessary delays for publications in the JASA, and how to review manuscripts for JASA.

9:25

4aID2. Realities of publishing in a journal: Why should you submit, what should you submit, and what problems might you encounter? Allan D. Pierce (Acoustical Society of America, Ste. 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502)

Attractions of a journal are its widespread availability, its archival (forever!) nature, its priority in literature searches, and its prestige. Articles should be readable to others in the field, be significant, and original. JASA is selective and imposes standards. Perceived quality is often measured by the "impact factor," which may have very little to do with the extent to which the journal fulfills the mission of the Society. Of great frustration to the editors is that a substantial fraction of authors who submit papers seem to be somewhat clueless as to what is a reasonable topic and what is a reasonable scope for an article in JASA. Frustrating also is that most good work reported at Society meetings never gets submitted. The present talk critically discusses the selection process, its peer-review aspects, and its flaws. The selection process deals with realities, such as that willing and competent reviewers are hard to find and that submitted reviews are often inane, with carefully selected highly expert associate editors who can make authoritative decisions without absolute reliance on external reviews. Suggestions are given on how to prepare a suitable manuscript and on how to cope with the vagaries of the peer-review process.

9:50

4aID3. Important aspects of editorial procedures for the information of inexperienced authors. Keith Attenborough (Dept. of Design, Development, Environment and Mater., The Open Univ., Milton Keynes MK7 6AA, UK, k.attenborough@open.ac.uk)

Several matters concerning the quality of papers and the editorial process have emerged from editorial activities for three acoustically-related journals (Applied Acoustics, Acta Acustica combined with Acustica, and the Journal of the Acoustical Society of America). Each of the journals suggests particular criteria to reviewers as the basis for their reviews. These factors are compared and discussed. Each of the journals offers various editorial decision options. These options and the uses that are made of them are discussed. Two common factors that influence editorial and reviewer judgements on submission quality include (a) the use of English and (b) the provision of a comprehensive and critical literature review. An increasingly important consideration is the availability of good reviewers in areas related to the topic of the paper. Sanitized examples illustrating some of these points are offered.

10:15—10:25 Break

10:25

4aID4. Preparing a submission to the Journal of the Acoustical Society of America on applied projects in architectural acoustics. Lily M. Wang (Architectural Engr. Prog., Peter Kiewit Inst., Univ. of Nebraska–Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681)

The Journal of the Acoustical Society of America (JASA) encourages authors to submit articles for publication that are based on more applied projects, such as those commonly found in the technical area of architectural acoustics. However, very few of this type have been published in the recent past. Suggestions on how such articles should be prepared and how they may meet the "significance" criterion will be given, from the viewpoint of a current JASA associate editor in architectural acoustics. Additionally a review of such articles that have been published in JASA within the past few decades will be provided.

10:50

4aID5. Manuscript preparation—How Acta Acustica united with Acustica compares with the Journal of the Acoustical Society of America. Dick Botteldooren (Dept. of Information Technol., Ghent Univ., Belgium, Dick.Botteldooren@intec.ugent.be) and Michael Vorlaender (RWTH Aachen Universivy, Germany)

Scientific writing is essential for academic progress and technical innovation. Rules and ethics for writing are well-defined, and detailed guidelines for authors are available. Nevertheless, journals have a specific flavor which is not only related to the focus of content. Thus, for authors preparing a manuscript, the choice of journal is an important issue. Based on survey material and analyses of past publications, differences between Acta Acustica united with Acustica (AA.A) and JASA will be highlighted. Journal policy has to deal with the variety of manuscript content covering the span from fundamental science to technical applied papers. The readers expect both, and this is even more important for society journals like JASA and AA.A. Review processes that guarantee equal quality of technical and purely academic contributions, yet keep in mind their own specificity, are accordingly difficult to set up. Recent changes in handling the review of technical and applied papers in AA.A will be discussed.

11:15—11:45 Panel Discussion

THURSDAY MORNING, 21 MAY 2009 STUDIO SUITE, 9:00 TO 11:15 A.M.

Session 4aMU

Musical Acoustics: Musical Perception and Modeling

Diana Deutsch, Chair
Dept. of Psychology, Univ. of California, San Diego, La Jolla, CA 92093-0109

Contributed Papers

9:00

4aMU1. Absolute pitch among students in an American music conservatory: Association with tone language fluency. Diana Deutsch, Kevin Dooley (Dept. of Psychol., Univ. of California, San Diego, La Jolla, CA 92093), Trevor Henthorn (Univ. of California, San Diego, La Jolla, CA 92093), and Brian Head (Univ. of Southern California, Los Angeles, CA 90089)

Absolute pitch (AP), the ability to name a musical note in the absence of a reference note, is extremely rare in the United States and Europe, and its genesis is unclear. The prevalence of AP was examined among students in an American music conservatory, as a function of age of onset of musical training, ethnicity, and fluency in speaking a tone language. Taking those of East Asian ethnicity, the performance level on a test of AP was significantly higher among those who spoke a tone language very fluently than among those who spoke a tone language fairly fluently, which was in turn higher than among those who were not fluent in speaking a tone language. The performance level of this last group did not differ significantly from that of Caucasian students who spoke only intonation language. An advantage to early onset of musical training was found, but did not interact with the effect of language. Further analyses showed that the results could not be explained by country of early music education. The findings support the hypothesis that the acquisition of AP by tone language speakers involves the same process as occurs in the acquisition of a second tone language.

9:15

4aMU2. The effect of musical experience on describing sounds with everyday words. Mihir Sarkar, Cyril Lan, Joseph Diaz, and Barry Vercoe (The Media Lab., Massachusetts Inst. of Technol., 20 Ames St., Cambridge, MA 02139, mihir@media.mit.edu)

Musicians often use non-technical words such as "warm," "sharp," or "sweet" to describe sound quality. Commonplace experience indicates that the descriptions of a sound by a diverse group of musicians may vary, suggesting that musical background may influence one's interpretation of a sound. A research study was carried out targeting 844 subjects of varying musical backgrounds where each subject had the chance to match various words to sound samples. Each subject was assigned to one or more musical categories (strings, woodwinds, electronic, percussion, brass) based on previous musical experience, and the results were compared across categories. Statistical measures were employed to determine if a correlation existed between musical background and survey responses. After analyzing the results from all sound-word combinations, it was determined that the musical background had no effect on the selection of words. Because of the nature of statistical hypothesis testing, the expected rate of false positives was greater than the proportion of statistically significant sound-word combinations. From the data collected in this user study, it is reasonable to suggest that the description of musical sounds is an innate skill that is not influenced by musical background and training.

9:30

4aMU3. A spectral analysis of the tuba mirum section of six live concert performances of Verdi's Requiem as conducted by Arturo Toscanini. Robert C. Chapman (11721 W. Brandt Ave., Littleton, CO 80127)

Arturo Toscanini conducted Verdi's Requiem 29 times with ten different orchestras between the years 1902 and 1951. Of these 29 performances, six of them were recorded between the years 1938 and 1951 with three different orchestras. An analysis of the Tuba Mirum section of this work was done to determine musical element consistency between different dates of performance and different orchestras. The Tuba Mirum section was selected because of the amount of control a conductor has to have over their musicians in order to obtain maximum musical effect. The musical elements considered in the analysis were rhythmical consistency, tempo, articulation, and dynamics. The amount of time to perform this section of music was also investigated.

9:45

4aMU4. Auditory effect of perturbing musical tones by interpolation with other musical tones. James W. Beauchamp (Dept. of Elec. & Comput. Eng. and School of Music, Univ. of Illinois at Urbana-Champaign, 1002 Eliot, Urbana, IL 61801, jwbeauch@uiuc.edu), Andrew B. Horner (Hong Kong Univ. of Sci. and Technol., Clear Water Bay, Kowloon, Hong Kong), and Richard H. Y. So (Hong Kong Univ. of Sci. and Technolgy, Clear Water Bay, Kowloon, Hong Kong)

The effect of blending variable amounts of a secondary tone with a primary tone by means of interpolating between their corresponding harmonic amplitudes was investigated. Original tones (bassoon, clarinet, flute, horn, oboe, alto saxophone, trumpet, and violin) were normalized with respect to fundamental frequency, duration, attack and decay times, and loudness; also, harmonic frequencies were flattened. However, the basic time variations of the tones' harmonic amplitudes were preserved. Interpolation was accomplished in the frequency domain using interpolation levels between 5 and 50%. While the effect of perturbation is highly dependent on the primary /secondary instrument pair, results show that the effect of a secondary instrument is heard most easily for primary instruments horn and bassoon and least easily for primary instruments trumpet and saxophone. On the other hand, clarinet and trumpet are heard most easily as secondary instruments, whereas bassoon and violin are slow to be heard. Discrimination scores were correlated with different spectrotemporal measures of the tone spectra: spectral incoherence, normalized spectral centroid deviation, and spectral irregularity. The only significant effect found was that primary instruments with high spectral incoherence tend to mitigate against perturbation by secondary instruments. [Work supported by Research Grants Council Grants 613806 and 613508.]

10:00—10:15 Break

10:15

4aMU5. Physical modeling of the piano: An investigation into the effect of string stiffness on the hammer string interaction. Charalampos Saitis (Computational Acoust. Modelling Lab., Music Technol./CIRMMT, Schulich Sch. of Music, McGill Univ., Montreal, QC, Canada, charalampos.saitis@mail.mcgill.ca), Sarah Orr, and Maarten van Walstijn (Queen's Univ. Belfast, Belfast, UK)

The stiff string wave equation has four solutions, two of which are fast-decaying waves introduced by the string stiffness. In the case of digital waveguide modeling of piano strings these are normally neglected. Some recent reports have suggested that all four traveling waves should be considered, at least at the neighborhood of interaction points (i.e., the hammer and the boundaries). This paper investigates the effect of omitting string stiffness in the context of sound synthesis of the piano by physical modeling. A stiff, lossy string with a spatially distributed hammer force excitation is implemented using both a finite-difference time-domain scheme and a digital waveguide model. The two models are designed so as to have the exact same features but for the two stiffness-related solutions. Numerical experiments are employed to study the contact force and string velocity signals for different initial hammer velocity values. The results generally confirm that the two fast-decaying waves have only a marginal effect on the overall string motion. However, small audible differences result for bass strings struck with high initial hammer velocities.

10:30

4aMU6. A real-time music synthesis engine for physical modeling of plucked string instruments. Huynh V. Luong (School of Comp. Eng. and Info. Tech., Univ. of Ulsan, Ulsan 680-749, Korea), Sangjin Cho, Jong-Myon Kim, and Uipil Chong (Univ. of Ulsan, Korea)

Music synthesis continues to offer huge potential possibilities for the creation of new musical instruments. One of the promising music synthesis techniques is physical modeling which produces output sounds that resemble much more closely their physical counterparts since it offers the potential of more intuitive control. However, this results in tremendous computational and I/O requirements, prohibiting real-time use in composition and live performance. To meet the performance requirements, we introduce a parallel processing engine. Whereas commercial digital signal processors such as TI TMS320C6x families use silicon area for large multiported register files, large caches, and deeply pipelined functional units, our parallel processing engine contain many more simple processing elements (PEs) for the same silicon area. As a result, our engine often employs thousands of PEs while possibly distributing and colocating PEs with the data I/O to minimize storage and data communication requirements. In this paper, we implemented physical modeling of a representative plucked Korean string instrument, called Gayageum, which has 12 silk strings. Our engine achieved 12-notes music synthesis in real-time at 44.1 kHz sampling rate for the physical modeling algorithms. This is in contrast to TI TMS320C6x, which achieves only single-note music synthesis. [Work supported by KOSEF, R01-2008-000-20493-0.]

10:45

4aMU7. Acoustical properties of pure sound piano wire. David Ripplinger, Brian Anderson, Tim Leishman, and William Strong (Dept. of Phys., Brigham Young Univ., Provo, UT 84602)

Pure sound piano wire is a stainless-steel wire that has only recently entered the U.S. market. Because of its different composition, which makes the wire more malleable, it should have a considerably lower amount of inharmonicity compared to regular steel wire. Measurements were conducted on several pianos with regular piano wire (Rosslau and Mapes) and pure sound in order to assess their differences in inharmonicity and tonal qualities. The pure sound wire produced measurably less inharmonicity than regular wire on the same kind of piano. However, this difference is much smaller compared to the difference between small and large pianos. This presentation will explain the methods of measurement and analysis, as well as simulations that were implemented in order to analyze the effect that inharmonicity has on the temperament, octave stretching, and interval patterns. [The Brigham Young University Department of Physics is acknowledged for the funding it provided for this research.]

11:00

4aMU8. Identification of impact sounds by professional percussionists. Ching-Ju Liu (Dept. of Communicative Disord., Univ. of Wisconsin, 1975 Willow Dr. Madison, WI 53706, chingjuliu@wisc.edu) and Robert Lutfi (Univ. of Wisconsin, Madison, WI 53706)

The present study was undertaken to determine the best of the ear's ability to identify the sounds of impact produced by the simplest of resonant objects (bars, plates, and membranes). We hypothesized that best performance would be achieved by professional percussionists who have had many years experience striking such objects to achieve desired nuances in sound. Five percussionists and five nonpercussionist musicians were recruited as participants from the University of Wisconsin School of Music. Additionally, 10 nonmusicians were recruited from the University at large. In a standard two-interval, forced-choice procedure, with and without feedback, listeners were asked to judge the impact sound corresponding to: (1) the greater force of impact; (2) the harder of two mallets; (3) the denser of the two objects struck; and (4) the point of mallet contact closest to the center of the object. Sounds were synthesized according to the equations of motion derived from a simple physical model which has been used in past studies [cf. Lutfi *et al.*, JASA 118, 393–404 (2005)]. Generally, the results provide little support for the hypothesis that professional percussionists are more adept than the rest of us at judging the properties of impact sounds. [Research supported by NIDCD.]

Session 4aNS

Noise, Architectural Acoustics and ASA Committee on Standards: Hospital Noise and Health Care Facilities

Erica E. Ryherd, Chair
Dept. of Mechanical Engineering, Georgia Inst. of Technology, Atlanta, GA 30332-0405

Invited Papers

9:00

4aNS1. The new comprehensive acoustical criteria for healthcare facilities. David M. Sykes and Gregory Tocci (Remington Partners, LLC, 23 Buckingham St., Cambridge, MA 02138)

New comprehensive acoustical guidelines for healthcare facilities were drafted for the American Institute of Architects and the American Hospital Association in 2005–6 by the 500-member TC-AA.NS.SC (Speech Privacy & Healthcare Acoustics) which is chaired by Gregory Tocci, David Sykes, and William Cavanaugh. Following two rounds of public review, these guidelines were approved for use by the Green Guide for Healthcare in 2007 and in early 2009 they were issued as the sole "Reference Standard" for two new LEED "Indoor Environmental Quality" credits. In early 2010, following two additional rounds of public review, the same guidelines will be released by the Facility Guidelines Institute and the American Society of Healthcare Architects (a division of the American Hospital Association) in the authoritative, code-level document, "FGI/ASHE Guidelines for the Design and Construction of Hospitals and Healthcare Facilities." Practitioners and researchers need to be aware of these criteria when designing acoustical solutions for healthcare facilities. This paper will present an overview of the guidelines. The speaker is co-chairman of the technical committee responsible for drafting the criteria and of ANSI S12 WG44, which is currently developing an additional new healthcare acoustics standard.

9:20

4aNS2. Comparing the sound environments in two critical care settings. Selen Okcu (GaTech, College of Architecture, Atlanta, GA 30332-0155), Erica Ryherd, and Craig Zimring (GaTech, Atlanta, GA 30332-0155)

Critical care nurses perform crucial tasks in complicated sound environments. The existence of many different noise sources (i.e., staff conversation, medical alarms, etc.) with different sound qualities cause nurses to experience constantly changing acoustic conditions while providing care for critically ill. Some of those acoustical qualities in critical care settings can negatively affect nurse well-being and task performance. In different critical care settings, the acoustic qualities can vary. Some architectural qualities of those settings can be an indicator of those variations. In this study, we documented the detailed objective and subjective sound environment of two critical care settings with different architectural layouts. The comparative analyses are used to understand the differences between two acoustic settings and the relationship between subjective and objective sound environments. Perceived qualities of the physical work environments are examined to explore the association between architectural features and acoustic qualities.

9:40

4aNS3. Evaluating the intensive care unit soundscape. Timothy Y. Hsu, Erica Ryherd (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405, gth776e@mail.gatech.edu), and Kerstin Persson Waye (Gothenburg Univ., 405 30 Gothenburg, Sweden)

The intermittent sounds in hospital wards may induce arousal among patients leading to responses such as annoyance, sleep disturbance, and cardio-vascular reactions. The sound environment as a whole may also affect the efficiency and general health among the staff. A series of studies are being conducted by the authors to evaluate the modern hospital soundscape including occupant response. Collaborations between engineering and medicine are being utilized to assess the soundscape from both a quantitative and qualitative standpoint. This talk will focus on soundscape evaluations of intensive care units. This includes a pilot study performed in a medical-surgical intensive care unit (ICU) of a Swedish hospital. Patients were monitored for 24 hours during their stay in the ICU and both acoustic and physiological data were simultaneously recorded. Additionally, the staff wore dosimeters and completed perception questionnaires. The methodology and analyses of these detailed acoustic measurements and preliminary subjective results will be discussed. [Work supported by ASA and Swedish FAS.]

10:00—10:15 Break

10:15

4aNS4. Mechanical equipment and ambulances. Strange bedfellows of cardiovascular services. Chad N. Himmel and Jack B. Evans (JEAcoust., 1705 W Koenig Ln., Austin, TX 78756, himmel@jeacoustics.com)

A case study will be presented regarding vibration and noise control design for an existing heart hospital. Facility expansion within a crowded site necessitates sandwiching a suite of new cardiac catheterization laboratories between mechanical equipment rooms above and the hospital's main ambulance driveway below. The facility expansion also includes cardiac computed tomography, magnetic reso-

nance imaging, and nuclear imaging, plus cardiovascular care, treatment, and administrative support spaces. Procedure and support spaces can be disturbed by mechanical equipment noise and vibration and noise from ambulance siren. Cardiac procedure instruments and imaging can be disturbed by low frequency noise and vibration from building mechanical systems, ambulance traffic, and building occupants. Design criteria for permissible building vibration and continuous background noise will be presented along with relevant sources and bases. Measurement evaluation results for design will be contrasted with criteria. Measures incorporated in design to mitigate vibration and noise on catheterization laboratories will be discussed. Measures considered in design include structural floor "detuning" concepts, stiffened supports for catheterization laboratory C-arms, mechanical equipment vibration isolation mounts, and floating concrete floor systems for mechanical rooms. If available, postconstruction spectral analysis measurements of building noise and vibration will be presented.

10:35

4aNS5. Predicting structure-borne noise for new construction adjacent to medical facilities. Richard A. Carman (Wilson, Ihrig & Assoc., Inc., 5776 Broadway, Oakland, CA 94618) and Gary M. Glickman (Wilson, Ihrig & Assoc., Inc., New York, NY 10004)

When new building construction is conducted in close proximity to existing medical or medical research facilities, structure-borne noise due to construction activities generating vibration can cause impacts to the occupants of the existing building. Theoretical modeling of structure-borne noise is very complicated and can be extremely time consuming. The feasibility of the two primary candidates (i.e., FEA and SEA) for modeling this phenomenon is questionable, at least at this point, for a number of reasons that are discussed. It is possible that the time will come when a theoretical method can be used reliably to predict structure-borne noise transmission. In the meantime, using an empirical approach seems to be the most viable. A case study is presented of a facility for medical research involving mice. Preconstruction projections of structure-borne noise and vibration are compared to those measured with an extensive monitoring program employed during construction.

10:55

4aNS6. Vibration sensitivity of optical microscopes in the healthcare setting. Hal Amick and Michael Gendreau (Colin Gordon & Assoc., 150 North Hill Dr., Ste. 15, Brisbane, CA 94005, Hal.Amick@colingordon.com)

The paper examines vibration criteria for benchtop and articulated floor-supported optical microscopes, comparing published specifications with data measured in hospitals for diagnostic purposes. A case study of vibrations that degraded orthopedic microsurgery is of particular interest. Revisions and enhancements of published criteria are proposed.

Contributed Paper

11:15

4aNS7. Soundscape analysis of a neonatal intensive care unit (NICU) facility. Gary Siebein (School of Architecture, Univ. of Florida, Gainesville, FL 32611-5702, gsiebein@siebeinacoustic.com), Reece Skelton, Victoria McCloud, Robert Lilkendey, and Hyun Paek (Siebein Assoc., Inc., Gainesville, FL 32607)

A soundscape study was made of an existing NICU facility in a major urban hospital to document the nature and magnitude of sounds experienced by newborn infants in this environment. Focus group discussions were held with hospital administrators, NICU staff, design team members, and families of patients to determine the types of sounds that are heard in normal operation of the NICU. Long-term average sound level measurements of general sounds in the NICU were made for several work shifts. Short-term acoustical measurements of specific acoustic events were also made to obtain octave band sound level data for each of the activities, medical equipment sounds, and building infrastructure sounds that comprise the soundscape of the unit. Audio and video recordings of the specific acoustical events that comprise the soundscape of the NICU were also made. The acoustical measurements and soundscape analysis are used to evaluate proposed acoustical design criteria for NICU facilities compared to nonacoustic health risks reported in the literature.

Session 4aPA

Physical Acoustics, Underwater Acoustics, and Engineering Acoustics: A Half-Century with the Parametric Acoustic Array I

Kenneth G. Foote, Cochair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Murray S. Korman, Cochair
Physics Dept., U. S. Naval Academy, Annapolis, MD 21402

Chair's Introduction—7:55

Invited Papers

8:00

4aPA1. Early years of the parametric array—An anecdotal history. David T. Blackstock (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029 and M.E. Dept., UT Austin, Austin, TX 78712-0292)

This review covers the period from 1960 to about the mid-1970s. Westervelt presented his theory of the parametric array at the June 1960 Acoustical Society Meeting, held at Brown Univ. In the same session, Bellin and Beyer reported experimental confirmation, primarily for the array in water. Despite the tantalizing properties of the parametric array—narrow low-frequency beam produced with a small source, absence of sidelobes, and potentially broadband operation—the acoustical community at first paid little attention. Berktay woke us up with a series of papers in the mid-to-late 1960s on possible applications. Interest in underwater parametric arrays then exploded, including attempts to make parametric receiving arrays practical. Definitive experimental confirmation of the parametric array in air was finally reported in 1973. However, practical applications of the airborne array did not appear until many years later.

8:20

4aPA2. Parametric acoustic arrays: A Bergen view. Halvor Hobæk (Dept. of Phys., Univ. of Bergen, Allegt. 55, N-5007 Bergen, Norway)

At the University of Bergen (UoB), Norway, research activity in physical acoustics started in the mid-1960s with investigations on the parametric acoustic array (PAA). The newly appointed professor in applied mathematics, Sigve Tjøtta, had some years earlier been at Brown University and was inspired by the concept at a fundamental level, but also wanted experimental confirmation. No previous acoustical activity existed at UoB. The PAA project was started as a master project at Department of Physics, where the main activity was in nuclear, high-energy, and ionospheric physics. Bellin and Beyer's experiment served as a model. The results provided new information on the axial and directional properties of the difference frequency wave field. Inspired by this, theoretical modeling continued along with further measurements. Other nonlinear effects like acoustic streaming (boundary layer, density gradient) were also investigated. In 1975, a project together with SIMRAD and Norwegian Technical University resulted in a bottom penetrating PAA, later commercialized as "TOPAS." Numerical modeling based on the KZK equation resulted in the "Bergen Code," still in use for computing nonlinear acoustic propagation problems. In later years activity at UoB has expanded to encompass linear physical acoustics of various sorts, occasionally using PAA as a tool.

8:40

4aPA3. Parametric acoustic array development at the U.S. Navy's New London, Connecticut laboratory. David G. Browning (139 Old North Rd., Kingston, RI 02881, decibeldb@aol.com), Mark B. Moffett (731 Annaquatucket Rd., North Kingstown, RI 02852), and William L. Konrad (54 Laurel Hill Dr., Niantic, CT 06357)

A brief history of the development of parametric acoustic sources at the U.S. Navy Underwater Sound Laboratory (USNUSL), and its successor, the Naval Underwater Systems Center is presented. Inspired by Robert Mellen, the Parametric Sonar Group was formed to explore the practical implications of Westervelt's idea for underwater acoustics. Spanning more than two decades, this research pursued various potential applications exploiting the unique characteristics of parametric sources, including echo-ranging in reverberant environments, communications involving voice, data, music, and video signals, target-strength measurements, bottom bathymetry, high-resolution sub-bottom profiling, and measurements of scattering and reflection from surfaces. Parametric source design procedures were developed, and guidelines were established for diagnosing and avoiding undesirable effects, such as difference-frequency instability and spurious nonlinearities, such as direct radiation from the projector, cavitation, and receiver nonlinearity. Theoretical models extended the Westervelt theory to realistic primary fields. For planar projector arrays, parametric source design curves, valid for levels in the primary far field, enabled designers to predict performance without resorting to computer calculations. To obtain source-levels and beam-widths within the primary near field, where the secondary signal is generated, a computer program, called CONVOL5 in its present configuration, can be used to obtain more detailed performance predictions.

9:00

4aPA4. Research on parametric arrays in Russia: Historical perspective. Lev A. Ostrovsky (Zel Technologies/Univ. of Colorado, Boulder, CO 80305)

This presentation is concerned with the Russian part of the history of parametric arrays (PAs). After the pioneering works of P. Westervelt in the United States and V. Zverev and A. Kalachev in Russia and a decade of relative isolation, Soviet and Russian researches became involved internationally in the work on the PAs and, more generally, of nonlinear acoustic beams. Several Russian institutions have been involved, including Moscow University, Andreev Institute of Acoustics in Moscow, Radiophysics Research Institute and Institute of Applied Physics in Gorky (now Nizhny Novgorod), Gorky University, Pacific Oceanological Institute in Vladivostok, and some others. Selected Russian contributions of different years are to be outlined in this talk, including: (1) The theory of nonlinear acoustic beams, e.g., the Khokhlov-Zabolotskaya- Kuznetsov equation and other models of nonlinear acoustic beams in application to the PA; (2) the limitation of the PA efficiency by shock wave formation; (3) possible enhancement of the PA radiation by bubble layers; (4) experiments in laboratory and in the ocean; and (5) some new developments such as generation of low-frequency shear waves by ultrasound beams with biophysical applications.

9:20

4aPA5. Reflections on the early days of parametric array research and development. Thomas Muir (Natl. Ctr. for Phys. Acoust., Univ. of Mississippi, One Coliseum Dr. University, MS 38677)

Following Westervelt's classic paper (1960/1963), there was a period of laboratory exploration, involving small-scale experiments and theoretical modeling, with Orhan Berktay's early work anticipating the significance of several practical applications. Robert Mellen organized the First International Symposium on Nonlinear Acoustics (ISNA) in 1968, which the preset author was privileged to attend. Research interest and momentum began to build, expanding to field and seagoing experiments, as well as computational modeling. Some of the seminal milestones and achievements of this period are identified and illustrated. These include theoretical work, which enabled new experiments, and vice versa, facilitated by research tool engineering. A number of classic problems were studied, including parametric transmission and the effects of finite amplitude attenuation. Applications to air, water, and sediment research evolved, including propagation and backscattering measurements, subbottom profiling, and modal effects and communication in shallow water. Parametric reception became a hot topic, along with its intricate signal processing. Focusing of finite amplitude beams was developed, involving the use of lenses, also utilized for steering and scanning parametric beams. The bandwidth capability of parametric interaction was recognized, as were parametric transients for echo-ranging. Imaging applications emerged for both harmonic and difference frequency radiations, and a spin-off to biomedical ultrasonics was forged. By the mid-1980s, a dozen ISNA symposia had been held and the field had blossomed.

9:40

4aPA6. Audio parametric arrays in air. Mark F. Hamilton (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713-8029)

Following Westervelt's discovery and theoretical description of the parametric acoustic array reported at the 1960 meeting of the ASA at Brown Univ., and the measurements in both water and air at ultrasonic frequencies, confirming Westervelt's predictions qualitatively, reported by Bellin and Beyer at that same meeting, it was not until 1973 that Bennett and Blackstock reported convincing quantitative comparisons of theory and experiment for a parametric array in air. While their experiments produced difference frequencies in the audio range, not until the late 1990s did application of the parametric array to directional audio transmission in air achieve any commercial success. This presentation will briefly review the history of developments leading up to modern-day audio parametric arrays. Challenges associated with transducer performance, signal processing, and acoustical measurements specific to audio parametric arrays in air will be described. Features of several audio parametric arrays currently on the market will be discussed. Finally, results will be presented from an ongoing collaboration between the author's group at Univ. of Texas at Austin and Professor Pierre Khuri-Yakub's group at Stanford Univ. to create audio parametric arrays in air using capacitive micromachined ultrasonic transducers to transmit the ultrasonic primary beams.

10:00—10:20 Break

10:20

4aPA7. High-powered parametric acoustic array in air. Robert W. Haupt (MIT Lincoln Lab., Active Optical Systems Group, 244 Wood St., Lexington, MA 02420, haupt@LL.mit.edu)

MIT Lincoln Laboratory has developed a prototype high-powered parametric acoustic array (HPPAA) for standoff acoustic excitation in several applications. Parametric arrays offer a highly directional, narrow beam mechanism to deliver sound in air to desired targets typically within a 100 m range. However, a difficult challenge arises in generating sufficient sound power at difference-frequencies below 1000 Hz at a range which can be critical for many target types. An important objective of the HPPAA design maximizes the difference-frequency pressure amplitude at range by maximizing the end-fire array length established by the PAA. The design optimizes the trade-offs between the three characteristic lengths that control the resultant end-fire array length including the pump or carrier wave attenuation, PAA aperture, and acoustic saturation of air. In field demonstrations, the HPPAA generated a carrier wave pressure power approximately 155 dB re-20-microPa one meter from the PAA face while generating a 300 Hz difference-frequency SPL of 90 dB re-20-microPa 8 meters from the PAA. These sound pressure levels at a few hundred hertz may enable safe standoff excitation and detection of buried landmines and may be useful in standoff nondestructive testing (NDT) damage detection and imaging of structures.

10:40

4aPA8. The development of a 75 kilohertz phase steered active parametric sonar system for subseabed target detection. Paul Lepper and Bryan Woodward (Dept. Electric and Elec. Eng., Loughborough Univ., Loughborough, LE11 3TU UK, p.a.lepper@lboro.ac.uk)

Parametric sonar systems offer a number of potential advantages over conventional sonar's that has encouraged investigation and implementation of nonlinear acoustic effects into sonar systems by numerous groups since Westervelt's discovery reported in 1956. The work represented here describes the development of an active parametric sonar system utilizing a 75 kHz primary frequency signal with measured secondary frequency generation from 1–22 kHz using a 2-D 729 element square array arranged into 13 staves of 54 elements each. The array, dedicated computer controlled signal synthesis and power amplifier systems were developed and tested during trials on the Fisheries Research Services trails site on Loch Duich, Scotland. Primary and secondary frequency source levels of 245 dB *re* 1 μPa-m and 195 dB *re* 1 μPa-m at 7 kHz were measured with −3 dB full angle beam-widths of 3 and 2.5°–7° for frequencies 3–11 kHz, respectively. Individual control of the thirteen stave (channels) allowed in-pulse scanning and beam steering across a 36° sector including demonstration of real-time dynamic stabilization in one plane. Results will be presented for implementation of the system into a towed body sonar used during sediment classifications trials off the coast of France and for target detection tasks in Scotland.

11:00

4aPA9. The parametric array in Berea sandstone: definitive experiments. Pierre-Yves Le Bas, James A. TenCate, Robert A. Guyer, and Paul A. Johnson (Geophys., MS D443, Los Alamos Natl. Lab., Los Alamos, NM 87545, pylb@lanl.gov)

Previous measurements of the characteristics of the parametric array in sandstone by Johnson and Shankland [J. Geophys. Res. **94**, 17729–17734 (1989)] were difficult to perform and only qualitative. Scanning laser vibrometers (Polytec) now make parametric array measurements in rock easier. However, hysteresis and memory effects play a strong role in the dynamic behavior of rocks and have the potential to mess up the creation of the parametric array in the medium. Thus, an experimental study was performed to find out just how well the "classical" theory of nonlinear acoustics works for a granular solid, a sandstone. An array of alternating PZT disk transducers was epoxied to a large block of Berea sandstone (1.2, 0.4, 0.4 m), every other one broadcasting with separate frequency generators and amps. Primary frequencies were around 100 kHz; difference frequencies of 10 to 20 kHz were observed. Two-D beampattern scans were taken and the results compared with calculations based on classical theory. The agreement is surprisingly good with a beta of around 200 and a Q (inverse attenuation) of 50.

11:20

4aPA10. Laboratory applications of truncated parametric arrays. Victor F. Humphrey (Inst. of Sound and Vib. Res., Univ. of Southampton, Southampton, SO17 1BJ, United Kingdom, vh@isvr.soton.ac.uk.)

Parametric arrays have characteristics that make them ideally suited for use as versatile sources in the laboratory for performing a range of measurements and acoustic studies. Such measurements can exploit the wideband character of the source to provide detailed acoustic response characteristics of systems as a function of frequency, while the short pulses and narrow beamwidths enable multiple reflections and unwanted echoes to be resolved or minimized. However, the need to make measurements in the nearfield of the array means that truncation of the nonlinear interaction region with an acoustic low-pass filter is advisable in order to generate a region free of secondary-sources and simplify the interpretation of the results. This truncation also avoids the possible influence of hydrophone nonlinearity. The effects of truncation on the array characteristics are illustrated using both model and experimental results. A number of examples of the use of arrays in the laboratory are given, including their use to study acoustic scattering from structures and diffraction from baffles, and measurement of reflection and transmission properties of materials. These demonstrate how the characteristics of parametric arrays can be exploited to make detailed, accurate measurements of relevance to underwater acoustics under laboratory conditions.

11:40

4aPA11. Parametric projectors protecting marine mammals from vessel collisions. Edmund R. Gerstein (Leviathan Legacy Inc., 1318 SW 14th St., Boca Raton, FL 33486, gerstein2@aol.com), Laura A. Gerstein (Leviathan Legacy Inc., Boca Raton, FL 33486), and Steven E. Forsythe (U.S. Naval Undersea Warfare Ctr., Newport, RI)

Marine mammals are vulnerable to ship collisions. Measurements of controlled ship passages through vertical hydrophone arrays demonstrate a confluence of propagation factors and near surface effects that obscure the sounds of approaching vessels which then pose serious detection challenges for marine mammals. Joe Blue, who first identified these challenges, later conceived of a parametric method to mitigate them. A highly directional, dual-frequency parametric array has been developed to reduce collision risks by selectively alerting only those animals in the direct path of approaching vessels. The system projector is comprised of multiple elements, band-centered to transmit a high carrier frequency along with a lower side band signal. A single-side band modulation and phase-shift method are employed. The non-linearity of water then demodulates the mixed high frequency carrier into a lower frequency waveform audible to both manatees and whales. The bow mounted array projects a narrow beam directly ahead of vessels, and "fills in" acoustical shadows to alert marine mammals of approaching danger. Controlled field tests of the manatee device in Florida's NASA wildlife refuge are proving effective. Real-world deployments on select Navy and DOD vessels are planned this year and sea tests of a larger whale system will start next year. [Funded by U.S. Department of Defense, Navy Legacy Resource Management Program.]

Session 4aPP

Psychological and Physiological Acoustics: Auditory Spatial Perception

G. Christopher Stecker, Chair
Dept. of Speech and Hearing Science, Univ. of Washington, Seattle, WA 98105

Contributed Papers

9:00

4aPP1. Revealing and quantifying temporal aspects of envelopes of high-frequency stimuli that lead to enhanced processing of interaural temporal disparities. Leslie R. Bernstein and Constantine Trahiotis (Depts. of Neurosci. and Surgery (Otolaryngol.), Univ. of Connecticut Health Ctr., Farmington, CT 06030)

This presentation concerns evaluating which aspects of envelopes of high-frequency waveforms foster efficient discrimination of interaural temporal disparities (ITDs). The experiments were designed to assess, quantitatively, the explanatory capability of metrics including: (1) the normalized fourth moment of the envelope (Y), (2) the normalized interaural correlation of the envelope, and (3) the dead time between peaks or maxima of the envelope. An ITD-discrimination study was conducted employing raised-sine stimuli centered at 4 kHz [John *et al.*, Ear and Hearing **23**, 106–117 (2002)]. The parameters of the stimuli that could be varied using such stimuli include: (1) the exponent of the raised sine, (2) the frequency of modulation, (3) the depth of modulation, (4) the degree of phase-scrambling applied to the spectral components of the envelope, and (5) the degree of phase-scrambling applied to the spectral components of the waveform. This set of parameters permitted us to construct an ensemble of stimuli that pit one metric against another in a manner that reveals the relative salience of particular temporal aspects of the envelope. Results will be discussed both in terms of the characteristics of the external, physical stimuli versus the stimuli as processed by the auditory periphery. [Research supported by research Grant Nos. NIH DC-04147 and DC-04073 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.]

9:15

4aPP2. Sensitivity to interaural level differences in high-rate high-frequency click trains: A functional magnetic resonance imaging assessment. G. Christopher Stecker and Susan A. McLaughlin (Dept. of Speech & Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105, cstecker@u.washington.edu)

While the auditory cortex (AC) is known to be necessary for accurate sound localization, the nature of spatial representation in AC remains poorly understood. Past animal studies have quantified sensitivity to auditory spatial cues [e.g., interaural level difference (ILD)] across subregions of AC. In this study, functional magnetic resonance imaging (fMRI) was used to assess ILD sensitivity across the surface of human AC. Normal-hearing listeners were presented with narrowband Gabor click trains (4000 Hz center frequency, 3 ms interclick interval) that varied parametrically in ILD over a range of +/− 30 dB. Simultaneously, whole-head fMRI images were acquired using a sparse imaging paradigm at 3 Tesla (32-slice EPI, 3.0×3.0×4.5 mm resolution, TR=12 s). During imaging, listeners detected rare target changes in ICI. Hemodynamic responses of compact AC subregions in individual subjects were assessed quantitatively using statistical pattern recognition with two goals: (1) to identify contiguous AC regions that share common patterns of sensitivity to ILD, and (2) to quantify, in information-theoretic terms, the ability of those responses to differentiate stimuli differing in ILD.

9:30

4aPP3. Precedence effect in children and adults: Effects of heard and unheard echoes on localization accuracy. Ruth Litovsky, Shelly Godar, and Phillip Wesolek (Waisman Ctr., Univ. of Wisconsin, 1500 Highland Ave., Madison, WI 53705, litovsky@waisman.wisc.edu)

The precedence effect is thought to facilitate sound localization in reverberant environments through localization dominance. This is demonstrated at short lead-lag delays, when a single fused auditory image is heard whose perceived location is dominated by the leading source. We studied localization dominance in children ages 4–5 and in adults. Stimuli were brief noise bursts, with lead-lag delays ranging from 1–100 ms; the lead location varied along the azimuth (−60, −40, −20, 20, 40, 60 deg) and the lag was fixed at 0 deg. Subjects indicated the perceived location(s) of each heard source. Localization accuracy was computed for the lead and lag at each delay. On a separate task, with delays varying randomly, subjects indicated whether they heard one or two auditory events, and echo threshold was obtained. Results show that localization accuracy for the lead declined as delays increased, suggesting that as the lag became more audible its interference in the localization process increased. For most subjects, the interference was predictable from independent measures of fusion echo thresholds. The impact of the lag, a single simulated echo, on localization accuracy, was greater for children than adults. These data suggest that the precedence effect is less effective at facilitating sound localization in young children than in adults. [Work supported by NIDCD Grant R01DC008957.]

9:45

4aPP4. Real-virtual equivalent auditory localization with head motion. Griffin D. Romigh and Douglas S. Brungart (Air Force Res. Lab., Wright-Patterson Air Force Base, OH 45433)

Several researchers have shown that individualized head related transfer functions (HRTFs) can be used to produce virtual sounds that are equivalent in terms of localization to free field sounds. Thus far, however, these results have only been shown in studies that have required listeners to keep their heads stationary during the playback of the virtual sounds. In this study, we investigated the performance limits of a virtual auditory display system that incorporated individualized HRTFs but allowed free head motion during the localization process. One key aspect of the system is a high-speed HRTF measurement process that allowed a full set of individualized HRTFs to be measured in less than 4 min. This made it possible to make an HRTF recording and complete a localization task using the resulting HRTFs within the same 30-min experimental session. The results show that equivalent free-field and virtual localization performance was achieved when the virtual sounds were generated in the same session using specially-designed open-ear headphones that did not need to be removed during the headphone equalization process. This indicates that equivalent real-virtual closed-loop localization is possible even with the truncated, interpolated, minimum-phase HRTF filters that are required for practical, real-world virtual audio systems.

10:00—10:30 Break

10:30

4aPP5. Head-related transfer function enhancement for improved vertical-polar localization. Douglas S. Brungart, Griffin D. Romigh, and Brian D. Simpson (AFRL/RHCB, Wright-Patterson Air Force Base, 2610 Seventh St., OH 45433)

Under ideal laboratory conditions, individualized head-related transfer functions (HRTFs) can produce virtual sound localization performance approaching the level achieved with real sound sources in the free field. However, in real-world applications of virtual audio, practical issues such as fit-refit variability in the headphone response and nonindividualized HRTFs generally lead to much worse localization performance, particularly in the up-down and front-back dimensions. Here we present a new technique that "enhances" the localizability of a virtual sound source by increasing the spectral contrast of the acoustic features that are relevant for spatial perception within a set of locations with nearly identical binaural cues (i.e., a "cone-of-confusion"). Validation experiments show that this enhancement technique can improve localization accuracy across a broad range of conditions, with as much as a 33% reduction in vertical-polar localization error for nonindividualized HRTFs measured on a KEMAR manikin; a 25% reduction in vertical-polar error for nonindividualized HRTFs measured on other human listeners; and a 33% reduction in vertical-polar error for individualized HRTFs presented under nonideal laboratory conditions (i.e., with headphone fit-refit variability). These results suggest that the proposed technique could provide benefits across a wide range of real-world virtual audio display applications. [Work sponsored by AFOSR.]

10:45

4aPP6. Azimuth dependency in auditory perception of speaker's facing angle. Hiroaki Kato, Hironori Takemoto, Ryouichi Nishimura, and Parham Mokhtari (NICT/ATR, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan, kato@atr.jp)

In pursuit of an ultimately realistic human-to-human telecommunication technology, the ability to auditorily perceive the facing direction of a human speaker was measured. A male speaker sat on a pivot chair in an anechoic chamber and spoke a short sentence (about 5 s) while facing either of eight azimuth angles (0=listener's direction, 45, 90, 135, 180, 225, 270, or 315 deg). These angles were set solely by turning the pivot chair. Blindfolded listeners heard the spoken sentence at a distance of either 1.2 or 2.4 m from the speaker and were asked to indicate the speaker's facing angle. The average of errors between listeners' responses and actual speaker's facing angles was 23.5 deg. Further analyses showed a significant effect of the speaker's facing azimuth on the accuracy of listeners' responses. Listeners made remarkably small errors (6.6 deg on average) when the speaker faced the listener's direction (azimuth = 0 deg). The second smallest errors were observed for the back angle (azimuth = 180 deg). Few front-back confusions were observed. This dependency of listeners' accuracy was extensively discussed from the interest of effective acoustic cues obtained by a precise computer simulation as well as an ecological perspective.

11:00

4aPP7. Strength and cardivascular fitness predict time-to-arrival perception of looming sounds. John G. Neuhoff, Katherine L. Long, and Rebecca C. Worthington (Dept. of Psych., The College of Wooster, Wooster, OH, jneuhoff@wooster.edu)

Perceiving rapidly approaching sounding objects can be critical for survival. In studies of "auditory looming" perception, listeners consistently perceive sound sources as closer than they actually are, resulting in an underestimation of arrival time (Neuhoff, Planisek, and Seifritz, 2009; Rosenblum, Carello, and Pastore, 1987). This effect has been argued to provide an evolutionary advantage by allowing more time to prepare for the source. However, critical to this argument is the timely engagement of motor behaviors. Here, we tested the hypothesis that listeners with lower levels of physical fitness would have a larger anticipatory bias in perceived auditory arrival time, and thus a larger margin of safety in response to looming sounds. Listeners judged the arrival time of a three-dimensional looming sound. Physical fitness was measured using recovering heart rate after exercise and grip strength. Results show that the anticipatory bias in perceiving looming sounds is negatively correlated with physical fitness ($r=-0.41$). Those least prepared physically to interact with a looming sound source have a greater perceptual margin of safety. The findings are consistent with an evolutionary explanation of the anticipatory bias for looming sounds and provide evidence for fitness-based perception-action links between the auditory and motor systems.

11:15

4aPP8. How visual cues help us understand speech in a complex environment. Lingqiang Kong and Barbara Shinn-Cunningham (Cognit. and Neural Systems, 677 Beacon St., Boston, MA 02215)

At a cocktail party, visual cues may help a listener by showing them *where* or *when* to direct attention, *what* acoustic modulations a target utterance contains, and/or *what* articulatory gestures produce the target. Here, we investigated target speech intelligibility while varying the visual cues available in a complex, confusing auditory scene. In all cases, subjects listened for a target utterance in the presence of multiple masker utterances with similar grammatical structure spoken by the same talker. The timing and direction of the target (and maskers) varied randomly, increasing the uncertainty about where and when to focus auditory attention. The number of correctly reported target key words measured performance. Performance tended to improve as the amount of visual information increased, particularly when masker phrases came from the direction of the target. Performance was generally similar whether listeners saw full videos of the target talker from the correct direction or only a static image of the talker at the right time in the correct direction. However, temporal information about where and when the target occurred improved performance over knowing only target location. Results suggest that in these scenes, visual cues aid target understanding by indicating where and roughly when to direct attention.

11:30

4aPP9. Stimulus continuity is not necessary for the salience of dynamic sound localization cues. Ewan A. Macpherson (School of Commun. Sci. and Disord. and Natl. Ctr. for Audiol., Univ. of Western ON, 1201 Western Rd., London, ON, Canada, N6G 1H1, ewan.macpherson@nca.uwo.ca)

Correspondence between head rotation and resulting changes in interaural difference cues provides information about sound source location. We assessed whether source continuity or merely relative displacement is necessary for use of this dynamic localization cue. Low-frequency (0.5–1 kHz) noise-band targets, not correctly localizable in the absence of head motion or for motion duration <50 ms [Macpherson and Kerr, APCAM (2008)], were presented while the listener performed a practiced head rotation at 50 deg/s. The stimuli were either continuous (a single burst gated on and off as the head entered and exited a variable-width spatial window) or discrete (two 20 ms endpoint bursts, triggered as the head entered and exited the window). Human listeners reported the apparent location of the stimulus by orienting with their heads subsequent to the initial head rotation. The minimum head movement angle (MHMA) necessary to resolve front/rear ambiguity was measured for each stimulus type. Similar MHMAs of 5–10 deg were measured for continuous and discrete stimuli, suggesting that endpoint "snapshots'' providing only displacement information are sufficient for use of dynamic localization cues. That parallels the finding that stimulus continuity does not improve detection of source motion [Chandler and Grantham, J. Acoust. Soc. Am. 106, 1956–1968 (1992)]. [Work supported by NSF and NIH/NIDCD.]

11:45

4aPP10. Localizing a speech target in a multitalker mixture. Norbert Kopčo (Dept. of Cybern. and AI, Tech. Univ. of Košice, Letná 9, Košice, Slovakia, kopco@bu.edu), Virginia Best, and Simon Carlile (Univ. of Sydney, Sydney, Australia)

Despite the importance of spatial hearing in everyday listening, little is known about the accuracy of sound localization in a complex mixture of sounds. Here we measured, for the frontal audio-visual horizon, how accu-

rately listeners could localize a female-voice target amidst four spatially distributed male-voice interferers in a moderately reverberant room. To examine whether listeners can make use of *a priori* knowledge about the configuration of the scene, we compared performance when the interferer locations were fixed (in one of five known patterns) to when the locations varied from trial to trial. The presence of interferers disrupted target localization, even after accounting for reduced target detectability. Randomizing the interferer locations had a moderate influence, degrading performance in some configurations but improving it in others. All effects were magnified when the target-to-interferer intensity ratio was reduced. The results confirm that spatial perception is disrupted by interfering sounds, and that this disruption is modified to some extent by listeners' expectations about the spatial arrangement of the scene. [Work supported by HFSP, NIH, VEGA, ARC and a Univ. of Sydney Postdoctoral Fellowship.]

THURSDAY MORNING, 21 MAY 2009 FORUM SUITE, 9:00 TO 10:00 A.M.

Session 4aSAa

Structural Acoustics and Vibration: Distinguished Lecture

Sean F. Wu, Chair
Dept. of Mechanical Engineering, Wayne State Univ., Detroit, MI 48202

Chair's Introduction—9:00

Invited Paper

9:05

4aSAa1. A residual-potential boundary for time-domain problems in computational acoustics. Thomas L. Geers (Dept. of Mech. Eng., Univ. of Colorado, Boulder, CO 80309, geers@spot.colorado.edu)

For many years, researchers have sought theoretically exact time-domain computational boundaries that are temporally, spatially, and geometrically local. Unfortunately, spatial locality and geometric locality cannot be simultaneously achieved for exact boundaries. *Absorbing* boundaries offer spatial locality but not geometric locality, whereas *impedance* boundaries offer the latter but not the former. Following a brief discussion of these topics, a new, theoretically exact impedance boundary is introduced that is based on modal *residual potentials* for the spherical geometry. The new boundary produces a set of first-order, uncoupled ODEs for *nodal* boundary responses and a set of low-order, uncoupled stepping formulas for *modal* boundary responses. The two sets are coupled through nodal-modal transformation based on the orthogonal surface functions for the spherical boundary. Numerical results generated with the boundary are presented for selected benchmark problems. Finally, extension of the method to other separable geometries for the wave equation is discussed.

THURSDAY MORNING, 21 MAY 2009 FORUM SUITE, 10:15 TO 11:55 A.M.

Session 4aSAb

Structural Acoustics and Vibration: Vibro-Acoustic Diagnosis and Prognosis of Complex Structures I

Wen Li, Chair
Dept. of Mechanical Engineering, Wayne State Univ., Detroit, MI 48202

Invited Paper

10:15

4aSAb1. Experimental validation of vibroacoustic reconstruction of a rectangular plate with different boundary conditions. Logesh Kumar Natarajan and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

This paper provides an experimental validation of vibroacoustic responses of a rectangular plate reconstructed using Helmholtz equation least squares (HELS) method. Experimental studies were conducted on baffled rectangular plates of different aspect ratios with free as well as clamped boundary conditions under random excitation. The radiated acoustic pressures were measured using a planar array of microphones at a very close distance to the plate surface and taken as input for the HELS codes. The normal surface velocity distributions were then reconstructed, and the results were compared against the benchmark data obtained using a laser vibrometer. The structural modes of the plate with different boundary conditions were compared with those obtained by experimental modal analysis. Good agreements were obtained for both clamped and free boundary conditions.

10:40

4aSAb2. Dynamic modeling of complex structures in a broad frequency range. Hongan Xu and Wen L. Li (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

A general method, the so-called finite substructurE method (FSEM), is presented for the dynamic analysis of complex structural systems. In this method, a complex structure is considered as a collection of a finite number of basic structural components such as beams, plates, and shells. Instead of seeking a numerical solution at a number of discrete or grid points, the current displacement solution is sought, over each component, as a continuous field in the form of Fourier series expansions. Thus, the number of degrees of freedom, which now represent the Fourier expansion coefficients, can be substantially reduced in comparison with a grid-based solution for the same spatial resolution. Mathematically, the resulting system tends to be better conditioned than those in the finite element methods as the number of DOF's increases with frequency. The robustness of this model for high frequency applications can be further improved by incorporating statistical processes into the modeling method to properly reflect the means or account for the uncertainties of certain input variables. The proposed substructure method is considerably different from the existing substructure techniques in that no modal data are required for any component. Numerical examples are presented to demonstrate the reliability of this method.

11:05

4aSAb3. Vibro-acoustic coupling of a rectangular cavity backed by a flexible panel with general boundary conditions. Jingtao Du, Zhigang Liu, Tiejun Yang (College of Power and Energy Eng., Harbin Eng. Univ., Harbin, 150001, People's Republic China), and Wen Li (Wayne State Univ., Detroit, MI 48202)

Vibro-acoustic coupling system composed of a rectangular cavity and a flexible panel is widely studied in the backgrounds of panel vibration analysis and active noise control in enclosed sound field either by sound or structural actuators, corresponding to ANC and ASAC, respectively. In the most of the current investigation, the boundary conditions of the flexible structure are limited to simple case, namely simply supported and/or clamped. As an important structural parameter, boundary condition has a great effect on the coupling characteristics of such vibro-acoustic system, and a good understanding on this phenomenon will be helpful to the system design as well as active noise control. In this presentation, the analytical model of a rectangular cavity backed by a flexible panel with general boundary conditions is developed. Two sets of improved 2-D and 3-D cosine series with supplementary terms are constructed to respectively describe the displacement and pressure distributions in the boundary structure and 3-D acoustic cavity. The aim of introduction of the supplementary terms is to overcome the potential discontinuity encountered along the structural boundary and the fluid–solid coupling interface respectively for the both fields mentioned above. Numerical calculations are carried out to show the effectiveness of the current method and to study the effect of the structure boundary on the coupling characteristics of the panel-cavity system.

11:30

4aSAb4. Semi-active shock control technique for two-stage vibration isolation system based on vibration absorption. Wanyou Li, Zhigang Liu, Xueguang Liu, and Xixia Chen (College of Power and Energy Eng., Harbin Eng. Univ., Harbin, 150001, People's Republic of China)

Active control of shock is more difficult due to its instantaneous nature. In this presentation, the use of vibration absorption is proposed to decrease the harm from shock. Simulation and experimental studies are carried out on a two-stage vibration isolation system. The dynamic model of the two-stage vibration isolation system with the semi-active vibration absorber installed is developed. The effect of various parameters of semi-active absorber is analyzed through numerical simulation. The correctness is also validated by the experimental studies. The results show that the optimal installed position is on the upper stage mass, and a set of the optimal absorber parameters exists to make the anti-shock effectiveness achieve best. Such best effectiveness is not affected by the shock amplitude, however, by the continuance time and the excitation waveform. A semi-active vibration reduction and shock resisting control system including electromagnetism semi-active vibration absorber, DSP controller, and constant current source; etc is also designed. The experimental studies on the control effectiveness and reaction time of such system are subsequently performed. It is demonstrated that such a system can work in two working conditions, namely, vibration reduction and shock resisting, these two statutes can transiently switch, and control the secondary shock response effectively.

Session 4aSCa

Speech Communication: Vowel Inherent Spectral Change

Geoffrey Stewart Morrison, Cochair
School of Language Studies, Australian National Univ., Canberra, ACT 0200, Australia

Peter F. Assman, Cochair
School of Behavioral and Brain Science, Univ. of Texas at Dallas, Richardson, TX 75083-0688

Chair's Introduction—8:00

Invited Papers

8:05

4aSCa1. Static and dynamic approaches to understanding vowel perception. James M. Hillenbrand (Dept. of Speech Pathol. & Audiol., Western Michigan Univ., Kalamazoo, MI 49008, james.hillenbrand@wmich.edu)

The purpose of this paper is to provide a broad overview of work leading up to the current view that vowel inherent spectral change (VISC) plays a significant role in the recognition of vowel identity. Although seldom if ever explicitly stated, the view that implicitly guided vowel perception research for many years assumed that nearly all of the information needed to specify vowel identity was to be found in a cross section of the vowel spectrum sampled at a reasonably steady portion of the vowel. There is now a considerable body of evidence showing that VISC plays an important role in the recognition of vowel identity. Evidence comes from: (1) measurement studies showing that many nominally monophthongal English vowels show significant spectral change throughout the course of the vowel; (2) statistical pattern recognition studies showing that vowel categories are separated with far greater accuracy by models that take spectral change into account; and (3) perceptual experiments showing that vowel steady-states are neither necessary nor sufficient for conveying vowel identity. In spite of this evidence, a static view of vowel identity continues to be implicitly assumed in many studies of vowel quality. [Work supported by NIH.]

8:25

4aSCa2. Spectral change in the front vowels of North American English. Terrance M. Nearey (Dept. of Linguist., Univ. of Alberta, Edmonton, AB T6G 2E7, Canada)

Assmann and Nearey [J. Acoust. Soc. Am. **80**, 1297–1308 (1986)] coined the term "vowel-inherent spectral change" (VISC) to refer to change in spectral properties inherent to the phonetic specification of vowels. Although VISC includes the relatively large formant movements associated with acknowledged diphthongs, it was explicitly extended to include reliable (but possibly more subtle) spectral change associated with vowel categories of North American English typically designated as monphthongs. This paper reviews statistical evidence of VISC in the formant patterns in front vowels of /hVd/ syllables in three regional dialects of English: Dallas, TX [Assmann and Katz, J. Acoust. Soc. Am. **108**, 1856–1866 (2000)], Western, MI [Hillenbrand *et al.*, J. Acoust. Soc. Am. **97**, 3099–3111 (1995)] and Northern, AB (new data). Results suggest that VISC patterns may be useful characteristics for assessing dialect differences. Evidence is presented for the importance of VISC in vowel perception, including recent evidence from our laboratories regarding perception by second language learners. A progress report is provided on research into how VISC is best characterized parametrically and which temporal regions of a vocoid may be most effective in summarizing VISC patterns in varying consonantal contexts.

8:45

4aSCa3. Formant-frequency trajectories as acoustic correlates to speech perception. Michael Kiefte, Tara Collins (School of Human Commun. Disord., Dalhousie Univ., 5599 Fenwick St., Halifax, NS B3H 1R2 Canada), Christian Stilp, and Keith R. Kluender (Univ. Wisconsin, Madison, WI 53706)

Formant trajectories are excellent vowel discriminants; within vowel, they are nearly constant across speaker size, age, and sex, and across consonantal contexts. However, this model assumes that formant peaks are perceptually important and that human listeners track formant-frequency changes across time. Speech-recognition applications have avoided formant frequencies due to the difficulty of reliable formant tracking. In addition, it is not actually known whether human listeners do indeed follow formants perceptually across time. This paper presents results from several studies that examine the relationship between changing formant frequencies and perception. Alternative perceptual representations of vowels, such as global spectral shape, are precluded by evidence that individual formant amplitudes are largely ignored in vowel perception. In addition, where other spectral properties appear to have a perceptual effect, it is because stimuli have used formants that do not change. When formants are changing, perceptual effects of spectral shape properties disappear. In terms of human formant tracking, perceptual extrapolation of a formant sweep is mostly dependent on peak frequency and not other properties related to spectral shape. This demonstrates that listeners do indeed follow formant-frequency changes as auditory objects. Further research on formant frequency perception will be described.

9:05

4aSCa4. Vowel-inherent spectral change enhances adaptive dispersion. Keith R. Kluender, Christian E. Stilp, Timothy T. Rogers (Dept. of Psych., Univ. of Wisconsin, 1202 W. Johnson St., Madison, WI 53706, krkluend@wisc.edu), and Michael Kiefte (Dalhousie Univ., Halifax, NS Canada, B3H 1R2)

Despite wide diversity among particular vowel sounds used across the world's languages, there are profound systematicities across languages. Whether sets of three, five, seven, or more vowel sounds are used, vowels that comprise these sets have substantial commonality across languages. Using static measures of vowel spectra, Lindblom and colleagues have demonstrated principles of adaptive dispersion through which the compositions of vowel inventories can be predicted on the basis of maximizing perceptual distinctiveness among the vowels within a language. Here, we address whether introduction of vowel-inherent spectral change is consistent with principles of optimizing perceptual distinctiveness between vowels. We find that vowel-inherent formant trajectories generally serve to further disperse vowel sounds across time. Trajectories of formants for vowel sounds that are relatively close in static measures (formant center frequencies: beginning, center, end) tend to be relatively distinct as measured by angles in F_1, F_2, F_3 coordinates. In a complementary fashion, vowels that share similar trajectories have relatively distinct static characteristics. This perceptual efficacy of vowel-inherent spectral change maintains across multiple place-of-articulation contexts. Across the vowel space and across consonantal contexts, vowel-inherent spectral change serves to increase adaptive dispersion and enhance perceptual distinctiveness. [Work supported by NIDCD.]

9:25

4aSCa5. Vowel-inherent spectral change and the second-language learner. Catherine L. Rogers and Merete M. Glasbrenner (Dept. of Comm. Sci. and Dis., Univ. of South Florida, 4202 E. Fowler Ave., PCD1017, Tampa, FL 33620)

Relatively few studies have directly examined the use of vowel-inherent spectral change by second-language learners, perhaps because it represents one of the subtler cues to vowel identity. Nevertheless, understanding non-native listeners' perception of this cue and its integration with other cues to vowel identity can be regarded as a method of investigating the mastery of cues to vowel perception that is needed for native-like perception. While several studies of second-language speech perception have demonstrated differences in cue use by second-language learners, our investigation of the use of dynamic spectral cues and temporal cues by native and non-native listeners suggests that relatively early learners of English as a second language do not appear to use the dynamic spectral cue differently from native English speakers when other cues are preserved. Instead, early learners' vowel perception appears to be less robust to removal of multiple cues. This apparent difference in even early learners' ability to use cues to vowel identity independently of one another or to change listening strategy when one cue is degraded may explain a portion of the increased challenge that even relatively early learners of a second language appear to experience in understanding speech in noisy environments. [Work supported by NIH.]

9:45

4aSCa6. Vowel inherent spectral change in forensic voice comparison. Geoffrey Stewart Morrison (School of Lang. Studies, Australian Natl. Univ., Canberra, ACT 0200, Australia, geoff.morrison@anu.edu.au)

Two-parameter models of vowel inherent spectral change, such as dual-target or target-plus-slope models, have been found to be adequate for vowel-phoneme identification. More sophisticated curve-fitting models do not appear to outperform such two-parameter models. This suggests that if only simple cues such as initial and final formant values are necessary for signaling phoneme identity, then speakers may have considerable freedom in the path taken between the initial and final formant values. If the constraints on formant trajectories are relatively lax with respect to vowel-phoneme identity, then with respect to speaker identity there may be considerable information contained in the details of formant trajectories. Differences in physiology and idiosyncracies in the use of motor commands may mean that different individuals consistently produce different formant trajectories between the beginning and end of the same vowel phoneme. If within-speaker variance is substantially smaller than between-speaker variance, then formant trajectories may be exploited for forensic voice comparison. This paper reviews a number of forensic-voice-comparison studies (including studies conducted using the likelihood-ratio framework) which have extracted information relevant to speaker identity from formant trajectories. For the purposes of forensic voice comparison, models using parametric curves are found to outperform simple two-parameter models.

Session 4aSCb

Speech Communication: Vowel Perception and Production (Poster Session)

Peter F. Assman, Cochair
School of Behavioral and Brain Science, Univ. of Texas at Dallas, Richardson, TX 75083-0688

Geoffrey Stewart Morrison, Cochair
School of Language Studies, Australian National Univ., Canberra, ACT 0200, Australia

Contributed Papers

All posters will be on display from 10:10 a.m. to 11:30 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 10:10 a.m. to 10:50 a.m. and contributors of even-numbered papers will be at their posters from 10:50 a.m. to 11:30 a.m.

4aSCb1. Developmental study of vowel-inherent spectral change. Peter F. Assmann (Univ. of Texas at Dallas, Richardson, TX 75083), Terrance M. Nearey (Univ. of Alberta, Edmonton, AB, T6G 2E7, Canada), and Sneha V. Bharadwaj (Univ. of Texas at Dallas, Richardson, TX 75083)

Children's speech differs from adult speech in several important ways. First, children have smaller larynges and supra-laryngeal vocal tracts than adults, with the result that their formants and fundamental frequencies are higher. Second, the temporal and spectral properties of children's speech are inherently more variable, a consequence of developmental changes in motor control. Both of these sources of variability raise interesting questions for theories of talker normalization and vowel specification. In the present study we compare the pattern of time-varying spectral change in vowels from a database of vowel recordings from adults and children ranging in age from 5 through 18 years from the Dallas, Texas region. Preliminary findings indicate systematic age-related differences in the average frequencies of the formants, but the pattern of vowel-inherent spectral change is well preserved across the age span investigated.

4aSCb2. Cross-generational differences in dynamic formant patterns in vowels. Robert Allen Fox and Ewa Jacewicz (Speech Percept. and Acoust. Labs., Speech and Hearing Sci. The Ohio State Univ., Columbus, OH 43210-1002, fox.2@osu.edu)

As the position of a vowel changes within the vowel space across generations of speakers, so does its dynamic formant pattern. This study examines variation in the dynamic patterns of vowel formants across two age groups: children (8–12 years) and older adults from their grandparents' generation (51–65 years). The cross-generational changes in vowels /ɪ, ɛ, æ/ were examined for each of the three regional variants of American English spoken in Southeastern Wisconsin (affected by the Northern Cities Shift), Western North Carolina (affected by the Southern Vowel Shift) and Central Ohio (not considered to be affected currently by any vowel shift). The following vowels in children's productions were monophthongized as compared to those of their grandparents' generation: Wisconsin /ɪ, ɛ/ (but not /æ/), North Carolina /ɛ, æ/ (but not /ɪ/), and Ohio /ɪ, ɛ, æ/. In addition to the reduced formant movement, some of the children's vowels had a different direction of formant change. These cross-generational changes were assessed in a set of measures including formant trajectory length, spectral rate of change and angle of formant change. The measures were calculated from formant frequencies extracted at points in the vowels corresponding to 20, 35, 50, 65, and 80% of the vowel's duration. [Work supported by NIH.]

4aSCb3. Generational and dialectal effects on children's vowel identification. Ewa Jacewicz and Robert Allen Fox (Speech Percept. and Acoust. Labs., Speech and Hearing Sci. The Ohio State Univ., Columbus, OH 43210-1002, jacewicz.1@osu.edu)

This study examines vowel identification by 8–13 years old children who grew up in either Southeastern Wisconsin (whose regional variant is affected by the Northern Cities Shift) or Western North Carolina (affected by the Southern Vowel Shift). In the first identification task, the children responded to words edited from sentences which elicited both stressed and unstressed vowel exemplars. This speech material was produced by multiple talkers representing two generations (children and older adults who represent their grandparents' generation). In the second identification task, the children were presented with citation-form tokens produced by three generations of talkers (children, their possible parents, and their possible greatgrandparents). Both within- and across-dialect vowel identification was examined. The cross-generational results showed that some vowels were identified more accurately when spoken by children, some when spoken by adults and for others there were no cross-generational differences. The cross-dialectal results indicated generally more accurate identifications of vowels produced by talkers from the same dialect region as the listeners. For selected vowels, there were significant interactions between dialect and generation. As a whole, the study shows children's sensitivity to cross-generational vowel changes and the attunement to their own dialect. [Work supported by NIH.]

4aSCb4. Vowel detection and vowel identification in long-term speech-shaped noise. Kathleen O'Brien, Ashley Woodall, and Chang Liu (Dept. of Commun. Sci. and Disord., 1 University Station A1100, The Univ. of Texas at Austin, Austin, TX 78712)

Psychometric functions of vowel detection and vowel identification were measured in long-term speech-shaped noise (LTSSN) for normal-hearing listeners. A four-interval forced-choice procedure was used to examine the accuracy of vowel detection in LTSSN with speech level presented from −10 to +5 dB sensational level relative to vowel detection thresholds obtained from Liu and Eddins' study [J. Acoust. Soc. Am. **123**, 4539–4546 (2008)]. The accuracy of vowel detection was significantly influenced by vowel category and sensational level. The threshold of vowel detection for each vowel and each listener was defined as the speech level at which 70.7% accuracy of vowel detection was reached. Vowel identification was then measured in LTSSN with vowel levels presented from 0 to 12 dB sensational level relative to individual thresholds of vowel detection, using a close-set 12-choice procedure. Results suggest that vowel identification was significantly affected by vowel category and sensational level. Altogether, the results of vowel detection and vowel identification indicate that, given the same signal-to-noise ratio, vowels are not equally audible and identifiable. Moreover, given the same sensational levels, vowels do not have the same identifiability, possibly due to the fact that some vowels have dominant confusing vowels while others do not.

4aSCb5. Auditory representation of vowel quality. Andrew B. Wallace and Sheila E. Blumstein (Dept. of Cognit. and Linguistic Sci., Brown Univ., Box 1978, Providence, RI 02909, andrew_wallace@brown.edu.)

It is generally assumed that the early stages of speech perception involve the extraction of some kind of generalized auditory patterns or properties from the peripheral input. The auditory representation that results is of con-

siderable interest, since it serves as the input to higher-level, speech specific processes of phonetic perception. The current research examines this auditory representation using a priming paradigm, in which perception of vowel targets is facilitated when the targets are preceded by acoustically matched nonspeech stimuli. By manipulating the acoustic parameters of these nonspeech "prime" tones, it is possible to determine the role of these parameters in the auditory stages of vowel processing. Previous results [Wallace and Blumstein, J. Acoust. Soc. Am. **119**, 3245 (2006)] suggest a short window of analysis of no more than 25 ms. In Experiment 1, frequency of nonspeech primes was varied, with results suggesting broad frequency tuning. In Experiment 2, primes matched to both F1 and F2 of the target vowels were found to elicit a greater priming effect than would be predicted by summing the response to separately presented F1 and F2 primes, suggesting that the auditory representation of vowels encodes combinations of formant frequencies.

4aSCb6. Cross-dialect differences in vowel identification. Amy T. Neel and Cai S. Ewing (Dept. of Speech & Hearing Sci., Univ. of New Mexico, Albuquerque, NM 87131, atneel@unm.edu)

In this study, identification of vowels produced by Michigan speakers by listeners of two dialects, Michigan (Inland North) and New Mexico (Western) was examined. New Mexico listeners identified sets of ten vowels (i, ɪ, e, ɛ, æ, ɑ, o, ʌ, ʊ, u) from 30 speakers in the Hillenbrand *et al.* (1995) database. Their results were compared to the results from the Michigan listeners in the original study. Preliminary results from ten New Mexico listeners show that overall identification rates across the 30 speakers did not differ significantly between dialects. However, NM and MI listeners differed on identification rates for individual speakers and particular vowels. They agreed on the worst speaker but differed by as much as 11 percentage points for others. NM listeners performed worse on /ɛ/ and /ɑ/ but slightly better on /ʌ/ than MI listeners. To determine the nature of cross-dialect differences in perception, acoustic data such as vowel space area, formant dynamics, and duration characteristics will be examined, and the relation of acoustic data to identification scores will be presented.

4aSCb7. Effects of dialect on vowel acoustics and intelligibility. Austin L. Oder (Dept. of Speech-Lang.-Hearing, Univ. of Kansas, Dole Ctr., 1000 Sunnyside Ave., Rm. 3001, Lawrence, KS 66045, aoder@ku.edu), Sarah Hargus Ferguson (Dept. of Speech-Lang.-Hearing: Sci. and Disord., Univ. of Kansas, Lawrence, KS 66045), and Cynthia G. Clopper (Dept. of Linguist., Ohio State Univ., Columbus, OH 43210)

A great deal of recent research has focused on the phonetic variation of American English vowels from different dialects. This body of research continues to grow as vowels periodically and unconsciously undergo formant movements that become characteristic of certain dialects (e.g., the Northern Cities Chain Shift and the Southern Vowel Shift). Two experiments using the Nationwide Speech Corpus [Clopper and Pisoni, Speech Communication, **48**, 633–644 (2006)] are exploring whether the Midland dialect is more closely related acoustically and perceptually to the Southern or to the Mid-Atlantic dialect. Experiment 1 consists of acoustic analyses of 11 English vowels from each of the three dialects. In Experiment 2, 11 vowels in /hVd/ format recorded from speakers of the three dialects are being presented to speakers of a Midland dialect for identification. This study will thus further our understanding of acoustic and perceptual differences between the most marked dialects (Mid-Atlantic and Southern) and the least marked dialect (Midland). [Work supported by a University of Kansas Honors Program Undergraduate Research Award.]

4aSCb8. Spectral analysis of the vowels of the Peking dialect. Wai-Sum Lee (Dept. of Chinese, Translation and Linguist., City Univ. of Hong Kong, 83 Tat Chee Ave., Kowloon, Hong Kong)

The paper presents the spectral characteristics of the dorsal and apical vowels in open syllables of the Peking dialect. Results show that (i) of the dorsal category the high and low vowels are monophthongs, whereas the mid-back vowels and rhotacized ə are diphthongal; (ii) the two apical vowels are similar in F1 and F2, but differ considerably in F3, with a lower F3 for the retroflexed than the plain one; (iii) the extensive overlaps between the vowel ellipses for the unrounded mid-back vowel and plain apical vowel in the F1F2 plane and between those for the rhotacized ə and retroflexed apical vowel indicate that the difference in perceptual quality between the paired vowels may be attributed to the dynamic property of the formant frequencies; (iv) as expected, the average F-values for the vowels are higher for females than males, however, the patterns of positions of the vowel ellipses in the F1F2 plane between the two genders are similar; and (v) the scaling between female and male F-values is nonuniform across vowel categories and across formant frequencies. The formant data will also be discussed in connection with the quality descriptions of the vowels in the past studies of the Peking dialect.

4aSCb9. Voice identification accuracy using multivariate vowel formant analysis. Al Yonovitz and Elvan Moss (Dept. of Commun. Sci. and Disord., Univ. of Montana, Missoula, MT 59812, al.yonovitz@umontana.edu)

Accurate and automated voice or speaker identification has been a major goal for those involved in forensic issues. In addition, voice and speaker identification has many applications in security and business. Numerous previous efforts to derive features for speaker classification have failed to achieve a sufficiently low error rate. In this study a linear discriminant analysis was independently performed for vowel formants (F1, F2, F3) for each of ten vowels. The standardized canonical discriminant coefficients (SCDC) weighted each of the formants within a database of voices. These SCDC values were then linearly combined to form a single scalar for each vowel. An analysis that considered the multivariate data vector composed of the ten vowel scalars was then used for classification. This algorithm was tested by comparing single voice exemplars to other voices in the database. Accuracy was extremely high using vowel formant values for classification.

4aSCb10. A duration-dependent account of coarticulation for hyper- and hypoarticulation. Harvey M. Sussman (Dept. of Linguist. and Commun. Sci. and Disord., Univ. of Texas, 1 University Station, Austin, TX 78712, sussman@mail.utexas.edu), Bjorn Lindblom (Stockholm Univ., Stockholm, S10691, Sweden), and Augustine Agwuele (Texas State Univ., San Marcos, TX 78666)

Hyperarticulated and hypoarticulated speech are accompanied by spectral expansion and spectral reduction of vowel nuclei, respectively. These shifts in F1/F2 vowel space directly affect degree of anticipatory coarticulation in consonant + vowel sequences apart from traditional vowel context effects. Examining the opposite conditions of emphatic stress [Lindblom *et al.*, J. Acoust. Soc. Am. **121**, 3802–3813 (2007)] and faster speaking rates (Agwuele *et al.*, Phonetica, in press) it was shown that coarticulatory effects could be documented independently of the expected vowel expansion /reduction effects. A modified locus equation regression metric was used in both studies to isolate alterations in F2 transition onsets due to prosodic and speech rate conditions apart from vowel space shifts per se. The current study provides a unified empirical and theoretical account for the opposite coarticulatory effects by providing duration data as the common variable tying both studies together. Articulatory and acoustic simulations of deeper (emphasis) and shallower (increased tempo) closures are provided to explain the shifts of observed F2 onsets relative to predicted F2 onsets due simply to expanded/reduced vowel space.

4aSCb11. Variability of vowel productions within and between days. Shannon L. M. Heald and Howard C. Nusbaum (Dept. of Psych., Univ. of Chicago, 5848 S. Univ. Ave., Chicago, IL 60637, smbowdre@uchicago.edu)

Theories of speech production and speech perception assume that phonetic categories can be characterized by stable properties. For example, the notion of stable vowel targets is used to organize articulation, and such targets could serve as the acoustic basis for recognizing vowel categories. However, there is substantial variation in the acoustic patterns of phonetic categories. The lack of invariance between acoustic properties and the phonetic categories of speech has posed a theoretical problem for understanding human speech perception. Although most theories vary in how acoustic-phonetic variability is approached, theories treat such variability as inherent in speech production or statistically regular or noise. In order to begin to understand acoustic-phonetic variability, we examined the naturally occurring variability in speech production over time. Participants' productions of seven different vowels ([IH], [EE], [EH], [AW], [AH], [AE], [UW]) were recorded in nine sessions: at three specific times in the course of each testing day (9 a.m., 3 p.m., 9 p.m.), every other day, over the course of 5 days.

Formant frequencies and variability of formant frequencies were analyzed to assess how vowel categories change over time. We will discuss the results and consider the theoretical implications of these results.

4aSCb12. Acoustic vowel space in pre-/r/ contexts: Shetland and American English. Peter Sundkvist (Dept. of English, Dalarna Univ., SE-79188 Falun, Sweden, psn@du.se)

The acoustic space available to vowel systems in pre-/r/ contexts commonly differs from that of other phonetic contexts. In English this space has gradually shrunk, which relates to changes in the phonetic nature of /r/, having gone from more consonantal articulations (tap, trill) to approximant and often complex articulations that have a stronger effect on the production and perception of adjacent vowels (e.g., "bunched" and retroflex /r/). This paper contains an acoustic study of pre-/r/ vowel systems in Shetland and American English. F1–F3 values were obtained from steady state vowel portions from words spoken in isolation. The acoustic vowel spaces and the positions of contrastive items within these are compared and discussed in relation to the phonetic nature of /r/ and its effects on vowel formants. In contrast to American English, Shetland maintains a trill or tap as the principal realization of /r/, and the full range of vowel contrasts found elsewhere is supported before /r/. As Shetland is one of the most conservative English dialects in this respect, the study may also offer some insight into the acoustic characteristics of English pre-/r/ vowel systems of earlier periods and the pressures affecting such systems.

4aSCb13. Relative contribution of jaw and tongue to the vowel height dimension in American English. D. H. Whalen, Aude Noiray, Khalil Iskarous, and Leandro Bolanos (Haskins Labs, 300 George St., St. 900, New Haven, CT 06511, whalen@yale.edu)

Vowels are typically described according to three articulatory dimensions: height, frontness, and rounding. Other researchers propose a role for the jaw in the height dimension. In the present study, we measured the relative contribution of the tongue and jaw for vocalic height distinction in American English vowels. Tongue and jaw motions were collected in six adult speakers for six vowels from all portions of the vowel space, embedded in [hVd] sequences. Tongue shape was captured via HOCUS (Haskins Optically Corrected Ultrasound System) which combines digital ultrasound imaging at 127 Hz with optical three-dimensional tracking of infrared emitting diodes (IREDs) positioned on the speaker's head and probe for subsequent decoupling of tongue motion from the jaw motion. Results showed a consistent role of the tongue in creating the vocalic constriction for the six subjects investigated, though with some idiosyncratic strategies. Jaw contribution was more sizable for vowel pairs traditionally contrasted by height than for tense/lax pairs, confirming previous findings from x-ray studies. One new result is the dominance of tongue contribution over the jaw's, even for vowels distinguished by height. [Work supported by NIH grant DC-02717.]

THURSDAY MORNING, 21 MAY 2009 PAVILION EAST, 8:00 TO 11:00 A.M.

Session 4aSP

Signal Processing in Acoustics: Pattern Recognition in Acoustic Signal Processing I

Grace A. Clark, Chair
Electronics Engineering, Lawrence Livermore National Lab., Livermore, CA 94550

Chair's Introduction—8:00

Invited Papers

8:05

4aSP1. Tutorial: Pattern recognition in acoustic signal processing. Mark Hasegawa-Johnson (Dept. of ECE, and Beckman Inst., Univ. of Illinois, 405 N Mathews, Urbana, IL 61801)

This tutorial presents a framework for understanding and comparing applications of pattern recognition in acoustic signal processing. Representative examples will be delimited by two binary features: (1) regression versus classification (inferred variables are continuous versus discrete); (2) instantaneous versus dynamic (inference algorithms consider only an instantaneous observation vector versus inference algorithms integrate observations with knowledge of system dynamics). (1. Regression) problems include imaging and sound source tracking using a device with unknown properties and inverse problems, e.g., articulatory estimation from speech audio. (2. Classification) problems include, e.g., classification of animal and human vocalizations and nonspeech audio events. Instantaneous classification and regression are performed using a universal approximator (neural network, Gaussian mixture, classification, and regression tree), regularized, if necessary, to reduce generalization error (resulting in a support vector machine, regularized neural net, pruned classification tree, or AdaBoost). Dynamic classification and regression are done by imposing a prior to characterize system dynamics. Depending on the prior, the resulting model may be called a hidden Markov model, finite state transducer, dynamic Bayesian network, or conditional random field (dynamic classification), or a Kalman filter, extended Kalman filter, or switching Kalman filter (dynamic regression).

8:45

4aSP2. Joint position/wave number and time/frequency features of signals. Leon Cohen (Dept. of Phys., Hunter College, 695 Park Ave., New York, NY 10021, leon.cohen@hunter.cuny.edu) and Patrick Loughlin (Univ. of Pittsburgh, Pittsburgh, PA 15261)

Joint position/wave number and time/frequency representations of nonstationary signals and noise yield features that may be used to classify signals. Among these are instantaneous frequency, group delay, instantaneous bandwidth, and other conditional, or local, moments. We present a review of these concepts and give examples from various fields. We also discuss how these features change when a pulse propagates through a dispersive medium and show that there are some moment features that are invariant to dispersion. We illustrate via simulation the utility of these dispersion-invariant features for classification of man-made objects (steel shells) in shallow water environments. Finally, we show how to transform equations of motion into phase space, and the advantages of this transformation, in terms of methods of approximation and characterization of signals and nonstationary noise. [Work supported by ONR.]

9:25

4aSP3. Testing for periodicity in speech waveforms. Betul Arda, Daniel Rudoy, and Patrick J. Wolfe (Statist. and Inf. Sci. Lab., Harvard Univ., Oxford St., Cambridge, MA 02138, patrick@seas.harvard.edu)

Speech waveform segments can roughly be categorized as voiced or unvoiced, in accordance with periodicity properties of the glottal source, and inferring this classification from data is in turn an important task underlying a variety of speech classification problems. This presentation describes a formal hypothesis testing framework for the detection of periodicity in general acoustic sources, with application to online voiced/unvoiced segmentation of speech signals. Beginning with the classical approach of Fisher, a variety of test statistics are proposed and analyzed in this context. Asymptotic analyses are provided, along with simulations to demonstrate the efficacy of such methods in the presence of compound periodicities, harmonic structure, and the low signal-to-noise-ratio environments typical of real-world speech applications. [Work supported in part by DARPA, NGA, and NSF.]

9:45

4aSP4. Entropy estimation using pattern matching in bioacoustic signals. John R. Buck (ECE Dept., Univ. of Massachusetts Dartmouth, 285 Old Westport Rd., N. Dartmouth, MA 02747, johnbuck@ieee.org) and Ryuji Suzuki (MIT, Cambridge, MA 02139)

Many bioacoustic signals consist of a sequence of discrete stereotyped sounds occurring in repeated patterns. A natural question to ask is how best to characterize the underlying structure of the source producing the sequence of sounds. The structure of the source manifests itself as constraints on the patterns observed in the sequence of sounds. These constraints determine how predictable the order of the sounds is. The information entropy of a discrete symbol sequence is a quantitative measure of how unpredictable the sequence is. A straightforward but biased technique for estimating the entropy of an unknown source is to substitute observed symbol frequencies into parametric models such as Markov models. More general nonparametric entropy estimators exploit the relationship between the entropy and the average length of matching patterns within the sequence. This nonparametric entropy estimate forms an upper bound on the amount of information conveyed by the sequence of sounds. Additionally, comparing entropy estimates from the parametric and nonparametric models provides a hypothesis test determining whether the parametric model sufficiently captures the constraints of the source. These techniques are illustrated in analyses of humpback whale songs and leopard seal calling bouts.

10:05—10:15 Break

Contributed Papers

10:15

4aSP5. Classification of audio signals using generalized spatial fuzzy clustering. Huynh V. Luong (Univ. of Ulsan, Korea, huynhlvd@yahoo.com), Cheol Hong Kim (Chonnam Natl. Univ., Korea), and Jong-Myon Kim (Univ. of Ulsan, Korea)

With the increasing use of multimedia data, the need for automatic classification and retrieval of certain kinds of audio data has become an important issue. In this paper, we propose an efficient method of audio signal segmentation and classification from audiovisual database. While conventional methods apply thresholding to audio features such as energy and zero-crossing rate to detect the boundaries, causing misclassification for audio signals which contain certain audio effects such as fade-in, fade-out, and cross-fade, the proposed algorithm, called general spatial fuzzy c-means algorithm (GSFCM), solves the problem by taking into account the local spatial information which is weighted correspondingly to neighbor elements based on their distance attributes. GSFCM detects the boundaries between two different audio signals, classifies segments, and then extracts unique feature vectors. This results in the accurate detection and classification. Experiment results for the audio signal from TV news program at 44.1 kHz with 30-min long confirm that the proposed method outperforms conventional methods in terms of accuracy of the audio signal classification. These results demonstrate that the proposed method is a suitable candidate for audio-video indexing which is compressed by MPEG. [Work supported by the MKE, Korea, under the ITRC supervised by IITA (IITA-2008-(C1090-0801-0039)).]

10:30

4aSP6. Detection and classification of defects in underground pipes using reflection coeffecients. Muhammad Tareq Bin Ali (School of Eng., Univ. of Bradford, BD7 1DP, UK, mtbinali@bradford.ac.uk)

Detection and classification of defects developed in underground pipes have been an important issue for a long time. A novel sensor has been developed to find defects (i.e., blockage, deformation, crack, etc.) in buried pipes. Theoretical analysis has been supported by the experimental results. The sensor consists of a speaker and four microphones. The location of the defect is determined by the reflected signal recorded in the microphones. The reflected signal from a particular defect is sent through filter banks to identify the signal properties, which are related to the property of the defect. Different types of defects have been simulated in the lab and in the field. The sensor has been found to locate and classify the defect considerably well.

10:45

4aSP7. Detecting and localizing pipe changes via matched field processing. A. Tolstoy (ATolstoy Sci., 1538 Hampton Hill Circle, McLean, VA 22101, atolstoy@ieee.org), K. Horoshenkov, M. T. Bin-Ali, and S. J. Tait (Univ. of Bradford, Bradford, UK)

In this work we shall demonstrate that the presence of some pipe changes, e.g., blockages can be detected via matched field processing given an array of two or more microphones. In particular, the location of such transient irregularities can also be determined using this technique. We shall show data processed with this technique for a variety of blockages in a variety of locations for a variety of pipes (concrete, PVC, and clay). There is no acoustic modeling of the fields required, and only the simple linear processor is applied. The method can be used to determine the location of change and its severity.

Session 4aUWa

Underwater Acoustics and Structural Acoustics and Vibration: Monostatic and Bistatic Detection of Elastic Objects Near Boundaries: Methodologies and Tradeoffs I

Mario Zampolli, Cochair
NATO Undersea Research Ctr., 19138 La Spezia, Italy

Karim G. Sabra, Cochair
Dept. of Mechanical Engineering, Georgia Inst. of Technology, Atlanta, GA 30332-0405

Chair's Introduction—7:30

Invited Papers

7:35

4aUWa1. Interpreting free-field scattering by metallic truncated cylinders and spherical shells. Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814, marston@wsu.edu) and Kyungmin Baik (Inst. of Sound and Vib. Res., Highfield, Southampton S017 1BJ, UK)

When investigating the complications to the scattering caused by proximity of a target to a boundary, a detailed ray-based interpretation of free-field scattering is often helpful. For tilted metallic cylinders having flat ends, when ka exceeds about 10, strong contributions to the backscattering have been identified with the radiation of sound by guided elastic waves associated with meridional rays, helical rays, and face-crossing rays [K. Gipson and P. L. Marston, J. Acoust. Soc. Am. **107**, 112–117 (2000)]. These features also affect the frequency response as a function of tilt and bistatic synthetic aperture sonar (SAS) and acoustic holographic images [K. Baik, Ph.D. thesis, WSU, 2008]. When considering the response of shells to transient insonification, it can also be helpful to construct the time-frequency response and to interpret the response using ray-based methods [D. H. Hughes, Ph.D. thesis, WSU, 1992; S. F. Morse and P. L. Marston, J. Acoust. Soc. Am. **111**, 1289–1294 (2002)]. Correct interpretation of the time-frequency response requires detailed consideration of the dispersion and attenuation of elastic waves guided by the shell and the shape of the radiated wavefronts. When considering low-frequency scattering, the inertia of the target significantly influences the scattering. [Work supported by ONR.]

7:55

4aUWa2. Boundary effects on elastic signatures of proud tilted aluminum cylinders. Jon La Follett and Philip L. Marston (Phys. and Astronomy Dept., Washington State Univ., Pullman, WA 99164-2814, lafollej@mail.wsu.edu)

Synthetic aperture sonar (SAS) and frequency response as a function of tilt (acoustic color) are two methods used to study target backscattering. Close proximity to a boundary can affect both the SAS and spectral signatures of a target. To improve understanding of these effects, scaled tank experiments were carried out on solid aluminum cylinders having flat ends and length to diam ratios of 2:1 and 5:1. To partially simulate the mechanisms present when an object is resting on the ocean bottom and illuminated at shallow grazing incidence, the cylinders were suspended through the air-water interface of a tank. Monostatic measurements were obtained as the source /receiver was scanned along a line parallel to the interface to produce SAS images. Backscattering measurements were also made as the target was rotated in a plane parallel to the interface, with the source/receiver stationary, to give the spectrum (color) as a function of tilt. Some of the elastic features in the SAS images and the acoustic color diagrams could be interpreted using a previously developed ray-based theory of generalized Rayleigh waves [K. Gipson and P.L. Marston, J. Acoust. Soc. Am. **106**, 1673–1689 (1999);, **107**, 112–117 (2000)]. [Research supported by ONR.]

8:15

4aUWa3. Monostatic and bistatic measurements of targets resting on or buried under the seafloor. Joseph Lopes (Naval Surface Warfare Ctr.-Panama City Division, Panama City, FL 32407-7001, joseph.l.lopes@navy.mil), Kevin Williams, Steve Kargl, Todd Hefner, Eric Thorsos (Univ. of Washington, Seattle, WA 98105), Philip Marston (Washington State Univ., Pullman, WA 99164), Iris Paustian, and Raymond Lim (Naval Surface Warfare Ctr.-Panama City Division, Panama City, FL 32407-7001)

Measurements were conducted to investigate monostatic and bistatic scattering of targets resting on or buried under a seafloor. The measurements were performed in NSWC-PCD's Facility 383, which is a 13.7 m deep, 110 m long, 80 m wide test pool with a 1.5 m layer of sand on the bottom. Two synthetic aperture sonar (SAS) rail systems were utilized in the measurements, and they were placed perpendicular to each other and oriented so as to look at the same region of the bottom. This strategy allowed a particular target and environment configuration to be set up and studied for both monostatic and bistatic geometries. Targets included a 0.61 m diam stainless steel shell and two 0.3 m diam solid aluminum cylinders with lengths of 0.61 and 1.52 m. This paper summarized the measurement setup and instrumentation. A sample of the results of these measurements will be presented. Subsequent papers will provide detailed analysis and comparisons with predictions of models. [Work supported by the Office of Naval Research and the Strategic Environmental Research and Development Program.]

8:35

4aUWa4. Measurement and modeling of solid cylinders near interfaces. Kevin Williams (College of Ocean and Fishery Sci., Univ. of Washington, 1013 NE 40 St., Seattle, WA 98105, williams@apl.washington.edu), Joe Lopes (Naval Surface Warfare Ctr.–Panama City, Panama City, FL 32407), Eric Thorsos (Univ. of Washington, Seattle, WA 98105), and Philip Marston (Washington State Univ., Pullman, WA 99164)

Acoustic scattering from solid cylinders located near interfaces include effects due to energy interacting with those interfaces. Therefore, modeling cylinder response also requires models of scattering from and penetration across those interfaces. The simplest modeling can be carried out using a plane wave approximation. Using this approximation finite element results for a solid cylinder in the freefield are used to calculate the acoustic scattering of the same cylinder located near an interface. These calculations are compared to experimental data for cylinder target strength and possible reasons for differences seen are discussed. The physical mechanisms responsible for the cylinder's response are examined and cylinder surface displacements are shown. [Work supported by the Office of Naval Research and the Strategic Environmental Research and Development Program.]

8:55

4aUWa5. Modeling the low- to mid-frequency scattering from a proud cylinder: Boundary and near-field effects. Mario Zampolli (NURC, Viale San Bartolomeo 400, 19126 La Spezia, Italy) and Kevin L. Williams (Appl. Phys. Lab., College of Ocean and Fishery Sci., Univ. of Washington, Seattle, WA 98105)

A 2:1 aspect ratio solid aluminum cylinder is placed on the planar interface between two fluids (water/sand and water/air), with a point source radiating at frequencies for which the acoustic wavelengths range from 0.5 to 12 cylinder lengths. The distance between the source and the target is approximately 100 wavelengths, relative to the center frequency, and a vertical receive array is placed near the source. The problem is studied using a finite-element target scattering model, which is capable of treating axially symmetric objects via the decomposition of the unknowns into a Fourier basis around the axis of symmetry. Since the axis of the cylinder is parallel to the planar interface, the overall problem is not axially symmetric. Nevertheless, an approximate solution is obtained, which takes into account the interface reflection of the incident field, as well as the first bounce of the scattered field (via the Helmholtz–Kirchhoff integral with layered medium Greens functions). Small variations in the source and receiver positions cause large changes in the received signal, which can be explained by a Lloyd mirror-like interference resulting from the coherent addition of the point-sources and image point-sources with which the incident field and the scattered field can be described.

9:15

4aUWa6. Acoustic color of elastic objects near boundaries: High-fidelity, high-speed, 3-D finite-element modeling. David S. Burnett (Naval Surface Warfare Ctr., Panama City Div., 110 Vernon Ave., Panama City, FL 32407, david.s.burnett@navy.mil)

NSWC PCD has developed a computer simulation system for modeling the acoustic color (target strength versus frequency and aspect angle) of realistic 3-D objects that are near to or straddling the interfaces between different fluids. It employs high-fidelity finite-element modeling of acoustic scattering from elastic objects (fully 3-D physics throughout object and environment), implemented in a custom-designed, scalable-architecture, multiblade rack system that efficiently manages the modeling of different parts of the frequency spectrum. The system automatically runs hundreds of thousands of finite-element models, dynamically changing the mesh resolution and outer fluid boundaries of the models as they sweep over frequency, and it produces a variety of outputs, the principal one being an acoustic color contour plot. This paper will begin with a brief overview of the acoustic color simulation system, followed by the results of two experimental validations of the system: (1) scattering from an aluminum cylinder in free space, and (2) scattering from an aluminum cylinder straddling the interface between two different fluids. The simulated acoustic color results will be compared with experimental, numerical, and theoretical results presented in other papers in this session. [Work supported by ONR and SERDP.]

9:35

4aUWa7. Modeling of acoustic scattering by sphere on a planar seabed. Zhengliang Cao, Shuanping Du, Shihong Zhou, and Fangyong Wang (Hangzhou Appl. Acoust. Res. Inst., Hangzhou 310012, China)

A model of acoustic scattering from spherical target above a planar seabed is advanced to a condition of irradiation by a point source, with both of T matrix method and complex images method. Compared to the other model [J. A. Fawcett and R. Lim, J. Acoust. Soc. Am. **114**, 1406–1415 (2003)], this model could be used to calculate scattering field from target above a planar interface in three dimension space or by bounded source beam. Comparing some coefficients by complex images method with that by analytical formula or numerical quadrature, the computing method of the model is examined to be efficient and accurate. In addition, numerical examples of a rigid sphere and an elastic spherical shell are compared for scattering field from the target on a fluid seabed, and the scattering field dependent angle is investigated from the different of grazing angle of incident wave. [Work supported by the National Natural Science Foundation of China (Grant No. 10704068).]

9:50—10:05 Break

10:05

4aUWa8. Three-dimensional structural-acoustics modeling and its validation for free-field and littoral environments. Saikat Dey (Global Strategies Group (North America), 2200 Defense Hwy., Ste. 405, Crofton, MD 21114, saikat.dey.ctr.in@nrl.navy.mil), Angie Sarkissian (Naval Res. Lab., Washington, DC 20375-5320), Eris S. Mestreau (Global Strategies Group (North America), Crofton, MD 21114), Brian H. Houston (Naval Research Lab., Washington, DC 20375-5320), and Larry Kraus (Global Strategies Group (North America), Crofton, MD 21114)

Accurate numerical modeling of structural acoustics scattering in littoral environments, in mid-to-high frequency regimes, presents several challenges, including the ability to model the truncation of the nonhomogeneous exterior domain with free surfaces. Additionally, the frequency-dependent nature of the problem introduces significant dispersion errors with increasing frequency. This makes low-order three-dimensional discretization prohibitively expensive for high accuracy computations. We present a modeling framework for three-dimensional acoustic scattering in littoral environments treated as two half-spaces defined as the fluid and the sediment, respectively. The fluid portion may terminate at a free-surface boundary. The sediment may be a damped fluid or poroviscoelastic solid. The model admits rigid or viscoelastic scatterer with complex three-dimensional shapes including internal structural details and fillings. Exterior domain truncation uses perfectly matched-layer (PML) approximations. An hp-finite-element approximation scheme is utilized to control dispersion errors providing high-accuracy solutions in mid-to-high frequency regimes. Large-scale three-dimensional problems consisting of millions of unknowns solved scalably using a combination of efficient multifrontal solvers and domain-decomposition (FETI-DP) techniques. Numerical results validating applications in free field as well as littoral environment against experimental data will be presented. [Work supported by HPCMP and ONR.]

10:20

4aUWa9. Coherent space-time-frequency processing to enhance the bistatic detection and classification of elastic shells in shallow water. Shaun D. Anderson and Karim G. Sabra (Woodruff School of Mech. Eng., Georgia Inst. of Technol., 771 Ferst Dr. NW, Atlanta, GA 30332-0405)

For underwater sonar, time-frequency analysis has been shown to be a powerful tool for detection and classification of man-made targets. For instance, with traditional monostatic systems, a key energetic feature of spherical shell is the coincidence pattern, or midfrequency enhancement, that results from antisymmetric Lamb-waves propagating around the circumference of the shell. The development by the Navy of mine countermeasure sonar systems, using a network of autonomous systems, provides a mean for multiple bistatic measurements, and thus potentially bistatic enhancement for target detection. However, time-frequency representations of bistatic simulations of scattered signals from spherical shell show that this coincidence pattern typically shifts in both time and frequency with respect to the monostatic case. Hence, this time-frequency shift is challenging for bistatic target detection algorithms based on standard array processing techniques. Design of robust multistatic sonar system based on the generalized space-time-frequency coherence of the bistatic measurements will be discussed. The influence of the source-receiver configuration and interface reflections on the proposed approach have been investigated numerically and experimentally using data collected in shallow water with an elastic spherical shell [Work supported by ONR Code321, N000140810087.]

10:35

4aUWa10. Experimental investigation of bottom target detection by single channel iterative time reversal. Yingzi Ying, Shengming Guo, Bingwen Sun, Li Ma (Inst. of Acoust., Chinese Acad. of Sci., 21 Beisihuanxi Rd., Beijing 100190, China, yingyz05@mails.gucas.ac.cn), Zhenghua Cai, Huamin Fu, and Bing Hu (Haisheng Technol. Ltd., Yichang 443003, China)

Iterative time reversal process will gradually lead echo waves to converge to a dominant narrowband resonant mode of the target and enhance the return level in noisy and reverberant environment. This technique is used in bottom target detection and an experiment has been performed in the Yellow Sea, China. The experiment is in a monostatic configuration, and the target, which is a 53 cm external diameter and 260 cm long stainless steel cylindrical shell with concrete interior, is resting on the seafloor, and the directional transceiver, which is a transmitter and receiver couple, is located right above the target. First, a broadband interrogation pulse is launched, and the echo is measured and a bandpass filter is applied to avoid transceiver response peak, then the signal is time reversed and retransmitted, and repeat above procedures iteratively. The bottom reverberation will gradually be suppressed, and the center frequency of converged signal corresponds to a target resonance frequency, which is different from inhomogeneous bottom response in no target case. The existence of target is determined by this important acoustic signature, and the results illustrate the feasibility of this method. [Work partially supported by the CAS Innovation Fund.]

Session 4aUWb

Underwater Acoustics and Signal Processing in Acoustics: Waveguide Invarient Principles for Active and Passive Sonars

Altan Turgut, Cochair
Acoustics Div. Code 7120, Naval Research Laboratory, Washington, DC 20375

Lisa M. Zurk, Cochair
Electrical and Computer Engineering Dept., Portland State Univ., Portland, OR 97207

Chair's Introduction—8:00

Invited Papers

8:05

4aUWb1. An overview of the waveguide invariant and some of its applications to signal processing. W. A. Kuperman, G. L. D'Spain, Peter Gerstoft, W. S. Hodgkiss, H. C. Song, and A. M. Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., U.C. San Diego, La Jolla, CA 92093-0238)

While the waveguide invariant is a compact descriptor of certain aspects of acoustic propagation in the ocean, it also has a broad range of applications from describing acoustic fluctuations to assisting in various types of signal processing. After presenting a brief theoretical and experimental overview of the invariant in terms of both normal modes and rays, some applications to signal and array processing are discussed.

8:35

4aUWb2. Effect of shallow water variability on the waveguide invariant distribution. Daniel Rouseff (Appl. Phys. Lab., College of Ocean and Fishery Sci. Univ. of Washington, Seattle, WA 98105) and Valery Petnikov (Russian Acad. of Sci., Moscow 119991, Russia)

When mapped versus range and frequency, acoustic intensity often displays a regular pattern of striations, ribbons of high intensity. The trajectory of these striations may be described by the waveguide invariant, commonly designated as beta. While beta may be formally an invariant quantity, it is not necessarily a constant, particularly in highly variable shallow water environments. To study fluctuations in beta due to fluctuations in the environment, it is useful to generalize from an invariant scalar to the waveguide invariant distribution. Through data analysis and simulation, the waveguide invariant distribution is calculated for different locations including the New Jersey shelf off the United States and the Kamchatka shelf off Russia. The effects of isotropic and anisotropic internal waves are quantified. [Work supported by ONR and the Russian Foundation for Basic Research.]

8:55

4aUWb3. Waveguide invariant motion compensation for adaptive beamforming. T.C. Yang and Altan Turgut (Naval Res. Lab., Acoust. Div., Washington, DC 20375)

Performance of high-resolution adaptive beamformers is significantly degraded when moving sources and interferers are considered. For a source changing bearing, the motion effect can be compensated by integrating over beam covariant matrix. For a source changing range, the motion effect can be compensated by a frequency change using waveguide invariant theory [T. C. Yang, J. Acoust. Soc. Am. **113**(1), 245–260 (2003)]. Simulated data showed that the signal beam of a moving source is narrow as for a stationary case when motion compensation is applied. Also, interference power has been suppressed by 6–10 dB when the motion of the interferer does not match that of the source. Motion compensation algorithms were also applied to ship noise data from the passing merchant ships, collected on a bottomed horizontal array during the RAGS 2003 experiment at the New Jersey Shelf. Similar results are obtained indicating narrower beamwidths and selective suppression of signals from the passing merchant ships when proper motion compensation is considered. [Work supported by ONR.]

9:15

4aUWb4. Passive striation based geo-acoustic inversion using ships of opportunity. Kevin D. Heaney (OASIS Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039, oceansound04@yahoo.com)

The coherent interaction of acoustic multipaths leads to an interference pattern in range. As range changes, the locations of the peaks (and troughs) of this interference pattern shift (as the group velocities of the modes change) in a sometimes predictable fashion. The intensity pattern of a moving source, plotted as a function of range and frequency, is commonly referred to as striation patterns. These patterns are commonly observed in the passing of a loud surface vessel in the spectrogram of a single hydrophone. In this paper, a geoacoustic method is presented based upon the quantitative measure of striation slope (wave guide invariant parameter) and the striation spacing in frequency (reciprocal time spread). Both of these parameters are generally sensitive to the geo-acoustic parameters of the sediment. In 2003 this technique was applied to a passing surface ship [K. Heaney, IEEE JOE **29**(1), 43–50 (2003)] and to a range-

dependent benchmark case [K. Heaney, IEEE JOE **29**(1), 88–99 (2003)]. The approach in these papers compared the measured acoustic observables with those from a table predicted by the forward computation using the broadband PE. Current implementation of the algorithm uses a Gaussian beam code to efficiently compute the acoustic observables.

9:35

4aUWb5. Waveguide invariance for active sonar. Jorge E. Quijano and Lisa M. Zurk (Elec. and Comput. Eng. Dept., Portland State Univ., 1900 SW 4th Ave., Ste. 160, Portland, OR 97201)

Active sonar signal processing in shallow water has proven to be a challenging problem due to the strong interaction of sound with the boundaries of the channel and the dependence on typically unknown environmental parameters. This has motivated research on properties of acoustic propagation that are not sensitive to those factors, such as the waveguide invariance. The invariance principle has found application in passive sonar signal processing by relating the frequency content of a broadband source to the range between source and receiver. More recently, experimental evidence has suggested that a similar structure exists for active sonar. This structure provides additional information about the location of a target, and information can be exploited in sonar processing algorithms such as target tracking. Data from several sea experiments have been analyzed to determine the behavior of the active invariance, and tank experiments have been designed to confirm the presence of the range-frequency structure in signals reflected by a moving target. This presentation provides an overview of the active invariance phenomena and describes the performance of a target tracking formulation that incorporates invariance structure into the state space representation for improved performance.

9:55

4aUWb6. Environmentally invariant features for classification of active sonar signals. Patrick Loughlin (Depts. of Bioengineering and ECE, Univ. of Pittsburgh, 745 Benedum Hall, Pittsburgh, PA 15261, loughlin@pitt.edu) and Greg Okopal (Univ. of Pittsburgh, Pittsburgh, PA 15261)

Dispersion and damping (frequency-dependent spreading and attenuation) can be significant in shallow water sound propagation. These propagation-induced effects can be detrimental to classification of active sonar returns because the observed backscatter depends not only on the target, but also on the propagation environment and how far the wave has traveled, resulting in increased variability in the received signals. We address this problem by extracting propagation-invariant features from the wave that can be used in an automatic classifier. In this talk, we review various moment-like features we have developed that are invariant per mode to dispersion and damping. Simulations of the backscatter from different steel shells propagating in a channel with dispersion and damping are presented to demonstrate the classification utility of the various invariant features. [Work supported by ONR Grant N00014-06-1-0009.]

10:15—10:30 Break

10:30

4aUWb7. Wavguide invariant-based characterization of wideband active sonar clutter discretes. Ryan Goldhahn, Jeffrey Krolik (Duke Univ., ECE Dept. Hudson Hall 130, P.O. Box 90291, Durham, NC 27708, rag15@ee.duke.edu), and Charles W. Holland (Penn State Univ. Appl. Res. Lab, Appl. Sci. Bldg., State College, PA 16804)

In active sonar, clutter discretes can produce strong, target-like returns which often produce false alarms of water column targets. While false alarm reduction methods based on statistical feature-based classifiers often lack sufficient training data, matched-field based classifiers often suffer from model mismatch. A waveguide invariant-based approach for estimating the magnitude short-time Fourier transform (STFT) of reverberation returns was presented [Goldhahn *et al.*, J. Acoust. Soc. Am. **124**(5), 2841–2851, (2008)]. In this paper, the waveguide invariant properties of the reverberation are used to predict the frequency selective fading of strong clutter discretes. In particular, comparison of waveguide invariant-based magnitude STFT estimates are compared with predictions made using a model of frequency-selective fading from a clutter discrete. The results are further compared with real sonar returns collected during the SCARAB98 experiment off a shipwreck in the Malta Plateau. The results suggest that when accurate characterization of the geoacoustic environment are available, the estimated STFT magnitude spectrum obtained by using a propagation and scattering model compare favorably with those obtained by averaging the reverberation along waveguide invariant striations. [Work supported by ONR.]

Contributed Papers

10:50

4aUWb8. Investigation of bistatic invariance principles for active sonars. Altan Turgut and Roger Gauss (Naval Res. Lab., Acoust. Div., Washington, DC 20375)

Application of bistatic invariance principles to mid-frequency active sonar systems is investigated using data collected during two recent experiments conducted at the Malta Plateau and East China Sea. Low-frequency (350–650 Hz) and mid-frequency (1.5–3.5 kHz) LFM pulses were transmitted and direct-blast and return signals from an echo-repeater and several strong scatterers simultaneously recorded on a towed array, a drifting volumetric array, and a moored vertical line array. At low frequencies, application of bistatic invariance principles to target detection is demonstrated. At mid-frequencies, the spectrograms of the direct-blast signals showed regular striation patterns that were used to estimate the waveguide invariant parameter beta. However, both measured and simulated spectrograms of the signals scattered from an oil rig indicated the complexity of the striation patterns. More complex patterns of striations in the measured spectrograms might be due to azimuthal dependency of the scattering kernel, source /receiver motion, and low SNR. [Work supported by the Office of Naval Research.]

11:05

4aUWb9. Full spectrum acoustic wave propagation prediction. Cathy Ann Clark (Naval Undersea Warfare Ctr., 1176 Howell St., B1320, Rm 457, Newport, RI 02841, cathy.clark@navy.mil) and Kevin B. Smith (Naval Postgrad. School, Monterey, CA 93943)

An overview of a normal mode solution to the Helmholtz wave equation to describe the underwater sound field for a fixed point source in a plane multilayered medium which utilizes Bessel functions of order 1/3 is presented. The mode functions are continuous across turning points of the separated depth-dependent differential equation due to careful selection of

the representations to be used for Bessel function arguments in various regions of the complex plane. A quotient involving vertical wave number and phase is seen to behave as a constant through turning points, enabling the mode amplitude functions to remain analytic, changing from oscillatory to exponential on traversing the turning point, thus enabling smooth incorporation of the continuous spectrum. The method also provides vertical directionality at all field points without post-processing the complex acoustic field. Comparisons of model results to a limited number of measured data sets and benchmark propagation codes are presented. Derivation and verification of the solution for bottom-interacting modes, including shear and compressional reflection and transmission for a layered bottom, as well as an extension into horizontally varying, shallow water environments are also discussed. Portions of this work have been published in the IEEE Journal of Oceanic Engineering.

11:20

4aUWb10. Acoustic ranging and waveguide invariant parameter estimation using virtual arrays. Altan Turgut (Naval Res. Lab., Acoust. Div., Washington, DC 20375)

A method for estimating the range of an unknown broadband acoustic source in a waveguide [Thode *et al.*, J. Acoust. Soc. Am. **108**(4), 1582–1594 (2000)] is revisited and extended to estimate both waveguide invariant parameter "beta" and source range in shallow water. In the new method, two or more vertical arrays are used without requiring a signal sample from a guide-source. It was shown that both methods are mathematically identical and they both provide robust range estimation even when the reference signal sample is used from different time and/or different vertical array location. It was also demonstrated that an image processing tool, Hough Transform method, provides robust parameter estimation from virtual array output data. In addition, the parameter estimation method was validated under both summer and winter conditions by using incoherent noise data to localize and track merchant vessels and to estimate waveguide invariant parameter. [Work supported by the Office of Naval Research.]

11:35

4aUWb11. Passive ranging using the waveguide invariant. Kevin L. Cockrell and Henrik Schmidt (Dept. of Mech. Eng., Massachusetts Inst. of Technol, Rm. 5-204, 77 Massachusetts Ave., Cambridge, MA 02140, cockrell@mit.edu)

A range versus frequency spectrogram of an acoustic field due to a fixed source in a waveguide will exhibit striations whose slopes depend on the range to the acoustic source and the value of the waveguide invariant. While many authors have pointed out that the range to an acoustic source can be estimated from the slopes of the striations in the spectrogram, few have presented an explicit algorithm to do so. An algorithm for estimating the range is presented and tested on experimental data collected in a shallow water waveguide during GLINT08, an exercise performed off of Pianosa Island, Italy. The experimental data consist of a fixed broadband acoustic source emitting energy at frequencies from 300 to 800 Hz, with a hydrophone measuring the acoustic field along a 1.75-km track directly away from the acoustic source.

THURSDAY AFTERNOON, 21 MAY 2009 GRAND BALLROOM II, 1:00 TO 3:30 P.M.

Session 4pAAa

Architectural Acoustics: Acoustics of Health and Healing Environments

Kenneth P. Roy, Chair
Innovation Ctr., Armstrong World Industries, 2500 Columbia Ave., Lancaster, PA 17604

Chair's Introduction—1:00

Invited Papers

1:05

4pAAa1. Acoustical designs in a new children's hospital. Francis Babineau, Jr. (Johns Manville, 10100 W. Ute Ave., Littleton, CO 80127, francis.babineau@jm.com)

The importance of noise control and acoustic comfort in healthcare facilities has been well documented. This issue is even more critical in pediatric facilities, given the often frightening and stressful circumstances associated with a child being in a hospital. Recently, The Children's Hospital in Denver, CO constructed a new facility in neighboring Aurora, CO. The new facility was opened in Oct. 2007, and one of the design goals was to improve the acoustic environment by implementing evidence-based design strategies. However, one of the challenges in improving hospital acoustics is to do so without introducing additional infection control risks. As part of the project, a series of noise measurements were performed at the old hospital and in analogous locations in the new hospital, after the new hospital was occupied. This paper will present the results of the noise measurements and discuss the impact (positive and negative) of various design elements on the acoustic environment.

1:25

4pAAa2. Perceptions and expectations of speech privacy in healthcare environments. Kenneth Good (Acoust. Privacy Enterprises, LLC, P.O. Box 252, Mount Joy, PA 17552) and Nikki Rineer (Hope Within, 4748 Harrisburg Pk., Elizabethtown, PA 17022)

Most methods for evaluating speech privacy were developed for offices and corporate environments and from the point of view of productivity and distraction impacts on the listeners. How do these methods translate to healthcare and other environments where confidential containment of information is required by law? This case study will look at the objective measurements of speech privacy along with patient subjective impressions and expectations surveyed.

1:45

4pAAa3. Evaluation and control of the acoustical environment in a long-term-care facility. Murray Hodgson and Gavin Steininger (Acoust. and Noise Res. Group, SOEH-MECH, Univ. of British Columbia, 3rd Fl., 2206 East Mall, Vancouver, BC, V6T1Z3 Canada)

This paper discusses the acoustical evaluation of the Minoru Residence long-term-care facility, to respond to concerns by its staff regarding the acoustical conditions. A review of existing acoustical standards with an analysis of their applicability to health-care facilities was conducted for the problems observed in the Minoru Residence. Measurements were made of the acoustical characteristics: unoccupied and occupied noise levels, reverberation times, Speech Intelligibility Index, and noise isolation. They showed that background noise levels in several key areas including the Rehabilitation Office and the Patient Lounges exceeded acceptable values. Reverberation times were excessive in the entrance lobby and patient common areas. The Speech Intelligibility values in the Nurse Stations and Rehabilitation Offices were below acceptable values. The noise isolation was inadequate between the entrance lobby and office areas. Recommendations were made for the improvement of the acoustical conditions. These recommendations include the reinforcement of the Front Office façade, and the application of acoustical ceiling tiles to the Rehabilitation Office and the entrance lobby.

2:05

4pAAa4. Achieving green design acoustical standards in healthcare facilities. Peter Holst and Ethan Salter (Charles M. Salter Assoc., 130 Sutter St., Ste. 500, San Francisco, CA 94104, peter.holst@cmsalter.com)

With the advent of the green guide for healthcare, and the introduction of LEED health care, which both include acoustical design credits, the benefits of good acoustical design have been recognized for improving patient recovery rates as well as staff health and efficiency. The acoustical issues in designing health care facilities cross the spectrum of acoustical design: sound isolation/speech privacy, room acoustics, sound masking, mechanical noise/vibration, environmental noise, and project-specific features, such as MRI and helipads. It is the responsibility of the acoustical consultant to address each item with the design team to develop cost-effective solutions to meet the criteria established by the green standards. This paper will provide background information on the acoustical design challenges for healthcare facilities and will discuss general approaches that achieve the goals of green acoustical design. Specific examples from projects will also be included, drawing from experience in design of hundreds of health-care facilities, as well as documentation of credits achieved for numerous medical centers that plan to submit for LEED healthcare. This paper will be a valuable resource for acoustical consultants in the design of all types of healthcare facilities, such as hospitals, central plants, medical office buildings, and nursing facilities.

2:25

4pAAa5. Safe and sound? Sleep disruption in healthcare facilities. Joanne Solet (Psychiatry, Cambridge Health Alliance, Cambridge, MA 02138, joanne_solet@hms.harvard.edu), Orfeu Buxton (Brigham & Women's Hospital, Boston, MA 02115), Andy Carballeira (Cavanaugh-Tocci Assoc, Sudbury, MA 01776), and Jeffrey Ellenbogen (Massachusetts General Hospital, Boston, MA)

National healthcare quality surveys have found that noise in hospitals is an urgent concern, showing negative impact on patient satisfaction. The purpose of this study was to develop sleep arousal probability threshold curves to specific hospital-based sounds, demonstrating their impact on all stages of human sleep. Recordings were captured of hospital sound sources corresponding to specific categories identified as salient in the American Institute of Architects' Draft Interim Guideline on Sound and Vibration in Healthcare Facilities. Fourteen sounds were calibrated for dynamic presentation through a speaker array at the sleep lab to deliver rising 5 decibel-step exposures from 40 to 70dB(A). Noise-related EEG arousals were quantified using current AASM criteria and summed for each sleep stage by sound type and decibel level to calculate arousal probability threshold curves. The tested stimuli evoked a range of arousal thresholds. At the 50% arousal probability level, stimuli spanned 15 dB(A) Leq in Stage 2 sleep, 17 dB(A) Leq in REM sleep, and 30 dB(A)Leq in Stage 3, the deepest sleep. The findings provide evidence that repeated arousals in all sleep stages occur even in healthy young adults when hospital sounds exceed 45dB(A); responses vary widely by stimulus types.

Contributed Papers

2:45

4pAAa6. Canadian hospital acoustical evaluation. Hind Sbihi and Murray Hodgson (Acoust. Res. Group, SOEH/Mech., UBC, 2206 East Mall, Vancouver, BC V6T1Z3 Canada, murray.hodgson@ubc.ca)

The aim of this study was to evaluate the acoustic conditions of two wards in a research/teaching hospital in British Columbia, Canada. The selection of the wards was based on managerial staff needs and perceptions of issues related to the acoustical working environment with respect to privacy and also aggressive behaviors. The two selected units were an adult emergency department and a long-term care facility where the patients population was a mix of elderly with different mental and physical health conditions. The methods will include long-term noise measurements, building acoustical measurements and interviews with nursing staff. In particular, measurements will be made of the following acoustical parameters in the facilities: unoccupied and occupied noise levels; reverberation time; Speech Intelligibility Index; noise isolation. Results will be evaluated by comparing them with acceptability criteria. Identification of nonoptimal aspects of the facilities acoustical environments will result from the consideration and analysis of staff responses and comparison with published guidelines.

3:00

4pAAa7. A comparison of sound transmission loss on metal stud partitions as the stud configuration changes. Aaron Betit (Veneklasen Assoc., 1711 16th St. Santa Monica, CA 90404)

The draft "Interim sound and vibration design guidelines for hospital and healthcare facilities" drafted by the Joint Subcommittee on Speech Privacy of the ASA recommends STC 50 partitions between exam rooms with no masking sound provided. These partitions are typically constructed of multiple layers of gypsum board on single steel studs, which are widely believed to achieve the required ratings based on published test reports. However, virtually all laboratory testing is with 25 gauge studs 24 in. on center, whereas with a 15 ft floor to floor height typical of hospitals, 16 in. gauge studs installed 16 in. on center are often required for structural reasons. There is little published data on the changes in sound transmission loss changes with stud gauge and spacing. A testing program was established, and transmission loss (STC) tests were performed on drywall partitions with various configurations of stud gauges, spacing, and layers of drywall. Measureable decrease in transmission loss as the studs become heavier and as the spacing between studs decreases was measured. The results of the testing program are presented.

3:15

4pAAa8. Noise inside a government owned hospital. Sergio Beristain (Acoust. Lab., E.S.I.M.E., IPN, IMA, P.O. Box 12-1022, Narvarte, Mexico, D. F. 03001, sberista@hotail.com)

A hospital is under evaluation in order to find out the most important noise sources. Noise measurements were carried out within and outside of the hospital, It was found that noise from the neighborhood was not an issue, but noises from within the hospital were loud enough to cause at least some disturbance to the patients and workers. Obviously the larger noise was the one coming from the machinery room, where the laundry, emergency power plant, etc., are located, but also in some of the most important care and treatment rooms, the noise generated inside was well over of the recommended limits for an installation of this type. Some measurement results are presented together with the description of the environment.

THURSDAY AFTERNOON, 21 MAY 2009 GALLERIA SOUTH, 2:00 TO 4:55 P.M.

Session 4pAAb

Architectural Acoustics, ASA Committee on Standards, and Noise: Indoor Noise Criteria

Lily M Wang, Chair
Architectural Engineering, Univ. of Nebraska--Lincoln, Omaha, NE 68182-0681

Chair's Introduction—2:00

Invited Papers

2:05

4pAAb1. Proposed components of an "ideal" indoor noise criteria rating system. Lily M. Wang (Architectural Engr. Prog., Univ. of Nebraska-Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681)

A number of studies have been conducted at the University of Nebraska on correlating human performance and perception to indoor noise criteria systems. Task performance and subjective perception data were gathered from subjects exposed to background noise conditions commonly due to mechanical systems, including some with tonal components and some with time-varying fluctuations. Results show that perceptions of annoyance and distraction are highly correlated to the character of the noise, but the current indoor noise criteria systems (such as noise criteria and room criteria) do not accurately reflect that relationship. This paper presents a proposal for what an indoor noise criteria rating system should ideally include, to quantify acceptable building mechanical system noise in commercial buildings. [Work supported by the American Society of Heating, Refrigeration and Air-Conditioning Engineers.]

2:30

4pAAb2. A case history comparing noise criteria. Jerry G. Lilly (5266 NW Village Park Dr., Issaquah, WA 98027)

A case history of an indoor HVAC noise problem in a new residential building will be presented. Noise measurements collected in the living room and in the bedroom of the impacted living unit will be examined using several of the available noise criteria methods including, NC, RC, NCB, and RNC (ANSI S12.2). Although the architect and the building owner believed that the HVAC noise was unacceptable, the measured noise levels met the NC, NCB, and RNC noise criteria. Only the RC method was able to accurately detect the problem.

2:55

4pAAb3. Evaluation of mechanical background noise outside the norm of generally accepted criteria. Andrew J. Boone and Michael R. Yantis (Sparling, 720 Olive Way, Ste. 1400, Seattle, WA 98101, aboone@sparling.com)

Cases are presented where mechanical background noise was found to be acceptable by clients, although it was above the threshold of conventional indoor noise criterion. Other instances are shown where noise levels fell within normal limits but were judged unacceptable. Examples include HVAC noise in offices and residences, chiller and other rooftop equipment noise in multifamily dwellings, and pump noise. Sound quality expectations and the perception of noise sources were found to play an important role in the evaluation of these noise sources.

3:20—3:35 Break

3:35

4pAAb4. Using indoor room criteria when the "room" is outside. Byron W. Harrison (TALASKE, 1033 South Blvd., Oak Park, IL 60302, byron@talaske.com)

Indoor noise criteria have applicability in the analysis of outdoor performance spaces. The presentation will provide an overview of the environmental noise issues at the Jay Pritzker Pavilion in Chicago, IL. The project design was largely influenced by noise concerns, in its physical form, audio design strategy, and building systems design. During the postconstruction tuning of the audio system a

number of parameters were adjusted to contend with the unusually low signal-to-noise ratio, including the overall loudness of amplified music, audio signal compression, the content and level of the acoustic enhancement system, and the approach to the active in-house mix of the audio system. An investigation was also undertaken to investigate the continuous and impulsive noise levels and frequency content at various times of day as compared with traditional environmental noise criteria and indoor room criteria. Subsequent studies regarding the impact of nearby construction noise on rehearsals and performances were influenced by the use of indoor noise criteria methods.

4:00

4pAAb5. The room noise criteria (RNC) metric. Paul Schomer (Schomer and Assoc., Inc., 2117 Robert Dr., Champaign, IL 61821, schomer@SchomerAndAssoc..com)

The recent ANSI S12.2:2008 room noise criteria contains both a survey and an engineering method to specify room noise criteria. The methods use A-weighting and extended NC, respectively. A new metric, titled like the standard, room noise criteria (RNC) is included as a diagnostic tool. It is based on human hearing and more correctly assesses low-frequency sound. In particular, it is sensitive to the standard deviation to random noise and/or low-frequency surging in the 16–125 Hz octave bands such as the sound that can be produced by HVAC systems or other devices. It provides a bridge between the NC and RC criteria by correctly predicting the need for the less stringent (at low frequencies) NC criteria when the HVAC system is well designed (no surging, moderate standard deviation) and also correctly predicting the more stringent (at low frequencies) RC criteria when the HVAC system noise has a large standard deviation and/or surging.

Contributed Papers

4:25

4pAAb6. Correlation of subjective and objective measures of spectral quality. Dakota M. Kelley and Lily M. Wang (Architectural Engr. Prog., Univ. of Nebraska—Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, dkelley@mail.unomaha.edu)

Common indoor noise criteria with spectral quality indicators have been compared to task performance and subjective noise perception. The analysis seeks to identify a relationship between criteria spectral quality ratings and human perception of common heating, ventilating, and air conditioning background noises. The three criteria evaluated include Balanced Noise Criteria, Room Criteria, and Room Criteria Mark II, due to their inclusion of rumble, hiss, or roar classifications. Study participants worked on typing, math, and verbal tasks while being exposed to various background noise signals, and then completed a questionnaire to describe their perception of the room acoustics. During the data analysis, background noise signals with a non-neutral spectral quality rating were weighted according to their respective spectral indicator, with greater weighting given to lower-frequency signals. Results demonstrate relationships between spectral quality designation and subjective perception of noise fluctuation and tonality. However, a relationship was not found between spectral quality designation and subjective perception of the same parameters (rumble, hiss, or roar). These findings warrant further investigation of the correlation between common criteria ratings and subjective perception. [Work supported by the Univ. of Nebraska Undergraduate Creative Activities and Research Experience Grant.]

4:40

4pAAb7. Acoustical criteria in a two-parameter system for evaluating impact noise insulation. John LoVerde and Wayland Dong (Veneklasen Assoc., 1711 16th St., Santa Monica, CA 90404, jloverde@veneklasen.com)

Experience indicates that impact noise complaints in multi-family joist-framed buildings fall into two broad classes: low frequency thudding from footfalls and mid- to high frequency noise from heel clicks, dragging furniture, etc. The authors have developed a two-parameter system for evaluating impact noise [LoVerde and Dong, J. Acoust. Soc. Am. **119**, 3220 (2006); **120**, 3206 (2006); **122**, 2954 (2007)] that offers considerable improvement over existing metrics (such as FIIC) in terms of both correlation with subjective reaction and comparison of materials intended for improving impact insulation. Based on this system, suggested criteria for impact noise levels are presented. The effects of various design parameters on noise levels are discussed.

Session 4pAB

Animal Bioacoustics: General Topics in Animal Bioacoustics II

David K. Mellinger, Chair
Oregon State Univ., Newport, OR 97365

Contributed Papers

3:00

4pAB1. Equine vocalizations: The start of a search for happiness. David A. Browning (Phys. Dept., Univ. of Rhode Island, 2 Lippitt Rd., Kingston, RI 02881, decibeldb@aol.com), Peter M. Scheifele (Univ. of Cincinnati, Cincinnati, OH 45267-0379), and Rebecca L. Pond (Univ. of Connecticut, Storrs, CT 06269)

As with all perissodactyls, the vocalizations of equines, specifically a horse's whinny, has a variable frequency (or melodic) component as well as just simple tonals. This appears to provide a primitive means of expression, simpler than any song or language but potentially more informative than the purely tonal moos or baahs of cattle or sheep (but a long way from the complexity of some birdsong). Sonograms are compared from in-barn whinnies recorded under apparently stressful (departure of a foal) and pleasant (arrival of the morning feed wagon) circumstances, with the same horse and various horses, to determine if distinctive patterns can be identified. We also compare these with greeting whinnies and departure whinnies. The ultimate goal is to be able to acoustically identify an expression of happiness.

3:15

4pAB2. A vocal repertoire of Asian elephants (*Elephas maximus*) and comparison of call classification methods. Sharon S. Glaeser (Dept. of Biology, Portland State Univ., P.O. Box 751, Portland, OR 97207, sharon@roguetechinc.com), Holger Klinck, David K. Mellinger (Oregon State Univ. and NOAA, Newport, OR 97365), Yao Ren (Marquette Univ., Milwaukee, WI 53201), Patrick J. Clemins (Arlington, VA 22203), Michael T. Johnson (Marquette Univ., Milwaukee, WI 53201), Mandy L. H. Cook, and Randy Zelick (Portland State Univ., Portland, OR 97207)

This study compares classification methods applied to an acoustic repertoire of the Asian elephant (*Elephas maximus*). Recordings were made of captive elephants at the Oregon Zoo in Portland, OR and of domesticated elephants in Thailand. Acoustic and behavioral data were collected in a variety of social contexts and environmental noise conditions. Calls were classified using three methods. First, calls were classified manually using perceptual aural cues plus visual inspection of spectrograms for differentiation of fundamental frequency contour, tonality, and duration. Second, a set of 29 acoustic features was measured for nonoverlapping calls using the MATLAB-based program Osprey, then principal component analysis was applied to reduce the feature set. A neural network was used for classification. Finally, hidden Markov models, commonly used for pattern recognition, were utilized to recognize call types using perceptually-weighted cepstral features as input. All manual and automated classification methods agreed on structural distinction of six basic call types (trumpets, squeaks, squeals, roars, rumbles, and barks), with two call types (squeaks and squeals) being highly variable. Given the consistency of results among the classification methods across geographically and socially disparate subject groups, we believe automated call detection could successfully be applied to acoustic monitoring of Asian elephants.

3:30

4pAB3. The hyena's laugh as a multi-informative signal. Nicolas Mathevon (ENES Lab, Univ. Jean Monnet, Saint-Etienne, France, mathevon@univ-st-etienne.fr) Aaron Koralek, Steve Glickman, and Frederic Theunissen (Berkeley, CA)

Many social mammals use vocalizations to encode information about sex, kinship, individual identity, and morphological cues, as well as motivational and physiological states. In spite of the importance of this multi-informative signaling for the maintenance of social groups, most investigations on information coding in vocal signals have focused on only one cue; e.g., individual identity. Using the opportunity of the captive colony of spotted hyenas *Crocuta crocuta* at the Field Station for the Study of Behavior, Ecology, and Reproduction (University of California, Berkeley), we recorded and analysed the hyena's giggle, one of the most well known calls of this large social African mammal. The acoustic analysis in both temporal and frequency domains was automated using a MATLAB customized routine. The fundamental frequency was tracked using two methods (cepstrum and autocorrelation) followed by a best guess using a Bayesian approach. The differences between giggles from different individuals or groups of individuals were assessed running a multiple analysis of variance (MANOVA in MATLAB), cross-validated by a permuted discriminant function analysis (pDFA, R software). The results show that the hyena's laugh encodes information about age, dominance status, and individual identity, giving to receivers some cues to assess the social position of an emitting individual.

3:45

4pAB4. An intelligent automated apparatus to assess absolute auditory thresholds in the laboratory mouse. Anna Pleuger (Charles Darwin Univ., Darwin, NT, Australia) and Al Yonovitz (The Univ. of Montana, Missoula, MT 59812)

The effectiveness of an intelligent behavioral training and testing apparatus was assessed by using this system to operantly condition mice. The method utilized an infrared grid that determined the location of the mouse and presented reinforcements for auditorily contingent bar-press behavior. The apparatus was fully automated. Absolute auditory thresholds were determined with the descending and ascending method of limits in the same group of C57BL/6 mice. There was a statistically significant difference between thresholds produced by these two methods, with the descending method producing more sensitive auditory thresholds. Thresholds were on average 4.4 dB lower and had smaller standard errors. Overall, the automated apparatus was a highly efficient and precise method for the operant conditioning of mice.

4:00

4pAB5. Social context influences acoustic communication in zebra finches. Clementine Vignal, Julie Elie (Univ. Jean Monnet, Saint-Etienne, France), Hedi Soula (INSERM U870, INSA, Lyon, France), and Nicolas Mathevon (Univ. Jean Monnet, Saint-Etienne, France)

During communication, a signal conveys information between an emitter and a receiver, but indirect receivers can eavesdrop on the interaction. In birds, communication has been demonstrated to often be under the influence of this eavesdropping. Social species show complex communication networks where audience drives individual behaviors. Zebra finches

(*Taeniopygia guttata*) are gregarious songbirds that live in social groups and form life-long pair-bonds. Previous studies showed that the vocal behavior of males highly depends on this audience effect. Males show mate calls preference over other female calls in the presence of an established male-female pair, but not in the presence of unmated male-female or male-male dyads. Males in social groups also show stronger vocal response to female calls than males in social isolation. In this study, we investigate whether female calls of varied social salience evoke differently male calling according to the audience. We show that social context modifies not only call rate in response to female calls of varied social salience, but also acoustic structure of evoked calls. Thus male distance calls are not stereotyped calls whose acoustic cues only convey bird's identity. We propose that fine spectral modifications of the calls could carry information about the emitter motivation.

4:15

4pAB6. Acoustic analyses of two undocumented sound patterns in the *Drosophila suzukii* and *D. takahashii* species subgroups. Yuwen Lai (Dept. of Linguist., Univ. of British Columbia, 2613 West Mall, Vancouver, BC V6T 1Z4, Canada), Shu-Dan Yeh (Stony Brook Univ., Stony Brook, NY 11794-5245), Jennifer Gleason (Univ. of Kansas, Lawrence, KS 66045), and John True (Stony Brook Univ., Stony Brook, NY 11794-5245)

Acoustic analyses of the courtship songs of the *suzukii* and *takahashii* subgroups in the *Drosophila melanogaster* species group were conducted. The primary and secondary pulse phases that are common in the *melanogaster* subgroup were not observed in these two subgroups. However, two undocumented sound patterns were discovered. *D. biarmipes* and *D. pulchrella* (*suzukii* subgroup) produce high amplitude, nonrhythmic "toot" sounds, which range from 82–158 ms in duration. The toot sound in *D. biarmipes* has a consistent dynamic frequency profile. It starts with an onset of 479 Hz and gradually falls to 422 Hz and then rises to 477 Hz. The toot sound in *D. pulchrella* has a significantly lower frequency when compared to *D. biarmipes*. Its frequency profile falls gradually from 352–259 Hz. In addition, a "turbo" sound was recorded in *D. prostipennis* (*takahashii* subgroup). It is composed of short, high frequency pulses (520 Hz) with 4 ms interpulse intervals. In the *melanogaster* subgroup, the parameters of pulses have been proposed to play an important role in female preference. The results of the present study suggest that there might be other parameters at play in the species investigated in the current study.

4:30

4pAB7. Recent insights on the mechanisms of frequency discrimination in cicadas (Hemiptera, Cicadoidea). Paulo J. Fonseca (Dept. de Biologia Animal and Centro de Biologia Ambiental, Faculdade de Ciências, Univ. de Lisboa, Bloco C2, Campo Grande, 1749-016 Lisboa, Portugal, pjfonseca@fc.ul.pt) and Axel Michelsen (Univ. of Southern Denmark, DK 5230 Odense M, Denmark)

Mate finding in cicadas is usually mediated by acoustic communication. Males produce a loud acoustic signal that is used to guide females toward singing males. The calling songs are frequently complex with changes in rhythm, amplitude modulation and, in many species, frequency modulation. Therefore, it is likely that the auditory organ encode some of those characteristics and that the nervous system may process the information and extract some species-specific parameters. The tympanic vibrations are transferred to the onion-shaped auditory organ, localized at some distance from the tympanum within the auditory capsule, through a stiff schlerotized apodeme. This configuration has raised problems to the understanding of how the different frequencies of the song, that Fonseca *et al.* (2000) have shown to be finely encoded at the level of the auditory interneurons in the cicada *Tettigetta josei*, are passed on to the auditory organ by the structures of the receptor. Using biophysical, electrophysiological, and anatomical measurements from the receptor structures of the cicada *Tettigetta josei*, a functional model that may allow for the above-mentioned frequency discrimination will be presented [Fonseca, P.J., Münch, D., and Hennig, R.M., "How cicadas interpret acoustic signals," Nature 405,297–298 (2000)].

4:45

4pAB8. Bayesian model-based technique for termites detection. Asif Mehmood (Dept. of Elec. Eng., Univ. of Mississippi, University Mississippi), Orwa Tahaineh, and John Seiner (Univ. of Mississippi, University, MS)

This paper presents a model-based approach to detect termites from their head banging acoustic signals, and is derived from Bayesian probability theory. The termite head banging is the loudest and most diagnostic sound that termites make, and can be utilized for termite detection. The laser Doppler vibrometry system is used to obtain the termite head-banging signals from infested wood. An algorithm based on Bayesian probability theory is developed to detect termites' presence. The atomic model that represents the termites' data is the sum of decaying sinusoidal signals. First the model selection is performed that tells us about the number of vibration frequency components present in the data under observation. Once the correct model is known, then the vibration frequency that corresponds to termites' head banging frequency is determined. The calculations are performed using the Markov chain Monte Carlo method. Monte Carlo integration is then used to approximate the marginal posterior probabilities for all the parameters, including the number of exponentials and whether a constant offset is present. The performance of this algorithm is evaluated by testing it on experimental data, and the results obtained reveal the excellent performance of the algorithm.

Session 4pBB

Biomedical Ultrasound/Bioresponse to Vibration: Cardiovascular Applications of Ultrasound Contrast Agents

John S. Allen, Chair
Dept. of Mechanical Engineering., Univ. of Hawaii, Honolulu, HI 96822

Chair's Introduction—1:15

Invited Papers

1:20

4pBB1. A new high frequency microultrasound system with applications in cardiovascular research. F. Stuart Foster (Sunnybrook Health Sci. Ctr., Univ. of Toronto, 2075 Bayview Ave., and VisualSonics Inc., 3080 Yonge St., Toronto, Canada)

The development of preclinical imaging using micro-MR, CT, PET, SPECT, optical, and ultrasound technologies has created a new paradigm for imaging in the laboratories of biomedical researchers. Once considered a luxury for isolated multiuser centers, microimaging platforms are now becoming mainstream in bioresearch where quantitative *in vivo* imaging measurements of biomarkers and other endpoints are becoming a requirement of these investigations. Microultrasound has come a long way from its inception in the mid-1990s. This paper will detail the progression of preclinical microultrasound from mechanical to array based imaging systems. The technology of high frequency array based ultrasound imaging will be reviewed including details on the transducers and beamformer used in the first commercially available system. Applications of this system in the areas of cancer and cardiovascular disease will be described. The development of high frequency microbubble contrast modes based on linear and nonlinear signal processing will be discussed with relevant examples including imaging of VEGFR-2 and CD31 expression in disease models. All experiments with animals were done under a protocol approved by the Sunnybrook or VisualSonics Animal Care Committees. [The author declares a significant financial interest in VisualSonics Inc.]

1:40

4pBB2. Identifying and controlling acoustic bioeffects. Robyn K. Schlicher (Dept. of Chem. and Biomolecular Eng., Georgia Inst. of Technol., 315 Ferst Dr., Atlanta, GA 30332, rschlicher@chbe.gatech.edu), Joshua D. Hutcheson, Daniel M. Hallow (Georgia Inst. of Technol., Atlanta, GA 30332), and Mark R. Prausnitz (Georgia Inst. of Technol., Atlanta, GA 30332)

Ultrasound exposure causes internalization of a wide variety of molecules in cells and tissues, especially in the presence of acoustic cavitation. However, high levels of uptake are often accompanied by cell death. This study first analyzed intracellular uptake and cell viability after exposure of porcine carotid arteries to ultrasound *ex vivo* by confocal microscopy and found that at moderate ultrasound pressure, there was extensive uptake by endothelial cells with little uptake by underlying smooth muscle cells, whereas at high ultrasound pressure there was extensive endothelial cell death and increase uptake by smooth muscle cells. To understand the mechanisms involved in cell uptake and death, sonicated DU145 cells were analyzed by high level microscopy, flow cytometry, and chemical analyses. This work showed that uptake was caused by transient wounds created in the cell membrane that resealed within minutes after sonication. It also identified and quantified seven different cellular responses to this wounding, including cell repair modes and four different death modes. To save cells from apoptotic death, which was found to be mediated by calcium, cells were exposed to a calcium chelator, which rescued approximately half of the apoptotic cells from death.

2:00

4pBB3. A predictive model using myocardial contrast echocardiography for patients presenting to the emergency department with chest pain and a nondiagnostic electrocardiogram. Sanjiv Kaul (Cardiovascular Div., Oregon Health and Sci. Univ., Portland, OR 97201)

Risk stratification of patients presenting to an emergency department (ED) with suspected cardiac chest pain (CP) and an undifferentiated electrocardiogram (ECG) is difficult. We hypothesized that a risk score incorporating clinical, ECG, Rand contrast echocardiography variables [regional function (RF) and myocardial perfusion (MP)] obtained at the bedside would accurately predict adverse events in occurring within 48 h of ED presentation. A logistic risk model was developed in the initial 1166 patients (cohort 1), and validated in another 720 patients (cohort 2). Any abnormality or ST-T changes on ECG (OR 2.5, 95% CI:1.4–4.5, $p=0.002$, and OR 2.9, 95% CI:1.7–4.8, $p=0.001$, respectively), abnormal RF with normal MP (OR 3.5, 95% CI:1.8–6.5, $p=0.001$), and abnormal RF with abnormal MP (OR 9.6, 95% CI:5.8–16.0, $p=0.001$) were found to be significant multivariate predictors of nonfatal myocardial infarction or cardiac death. The estimate of the probability of concordance for the risk model was 0.82 for cohort 1. Likewise, in cohort 2, the c-index for the risk model was 0.83. In conclusion, a model based on variables obtained at the patient's bedside can be used to accurately risk stratify patients presenting to the ED with suspected cardiac CP and a nondiagnostic ECG. Its application could enhance care of CP patients in the ED.

2:20

4pBB4. Diagnostic ultrasound combined with targeted microbubbles improves recovery following acute coronary thrombosis. Evan Unger, Terry Matsunaga (Dept. of Radiology, Univ. of Arizona, 1501 N. Campbell Ave., Tucson, AZ 85724), Feng Xie, and Thomas Porter (Univ. of Nebraska Medical Ctr., Omaha, NE 68198)

Forty-five pigs with acute coronary artery occlusions—low MI pulse sequence (CPS) to guide high MI (1.9 MI) pulses during infusion of platelet-targeted MBs or non-targeted MBs. Third group received no ultrasound/MB. All groups received pro-urokinase, heparin, and aspirin. Angiographic recanalization rates, resolution of ST elevation, and wall thickening were analyzed. Pigs receiving MB had more rapid replenishment of risk area (RA) (80% versus 40% for MB; $p=0.03$) and higher epicardial recanalization rates (53% versus 7% for pro-urokinase alone; $p=0.01$). Replenishment of contrast within RA with MB showed higher recanalization rates and higher rates of ST resolution (82% versus 21% for pro-urokinase alone; $p=0.006$). ST resolution occurred in six pigs (40%) with MB who did not have epicardial recanalization; five had recovery of wall thickening. Conclusions: IV MB with brief high MI DUS guided by CPS improves both epicardial recanalization rates and microvascular recovery.

2:40

4pBB5. Delivery of drug, or gene, to blood vessel wall using intravascular ultrasound and microbubbles. John Hossack (Biomed. Eng., Univ. of Virginia, Charlottesville, VA 22908-0759)

Atherosclerotic arteries are routinely treated using balloon angioplasty followed by stent placement. We developed a concept for a new therapy to prevent restenosis involving using ultrasound and microbubble-based drug/gene carriers. A number of early *in vitro* and *in vivo* results are presented. In particular, data for delivery of antiproliferative gene via microbubbles ruptured via catheter-based intravascular ultrasound (IVUS) at the site of vessel injury *in vivo* in a pig coronary artery are presented. Optimal design of a modified IVUS catheter is discussed. The proposed catheter includes a bubble port, an elongated single element transducer to provide radiation force to cause the bubbles to traverse to the vessel wall, and a bubble rupture transducer. The bubble rupture transducer is ideally coincident with an imaging IVUS scanning single element or annular phased array. These requirements provide the impetus to develop enhanced designs of both PZT based transducers and silicon micromachined transducers. In our initial pig study, transfection efficiency was measured using fluorescence microscopy and quantified as the percent of vessel perimeter cells expressing red fluorescent protein. We observed $23.3\pm6.0\%$ transfection in the treated vessel whereas the control exhibited $3.6\pm2\%$ transfection.

3:00—3:20 Break

3:20

4pBB6. Direct numerical simulations of bubble collapse near a tissue surface with the ghost fluid method. Hiroyuki Takahira and Kazumichi Kobayashi (Dept. of Mech. Eng., Osaka Prefecture Univ., 1-1 Gakuencho, Naka-ku, Sakai, Osaka 599-8531, Japan, takahira@me.osakafu-u.ac.jp)

The collapse of an air bubble induced by the interaction of an incident shock wave with the bubble near a gelatin or bone surface is investigated by using an improved Ghost Fluid Method (GFM). The motions of three phases for air inside the bubble, water, and gelatin (or bone) are solved directly by coupling the GFM with the level set method. The results show that the strong shock waves are generated not only when the bubble rebounds but also when the liquid-jet impacts the downstream surface of the bubble; the shock waves result in the depression of the gelatin (or bone) surface. Also, the penetration of the bubble into the depression of the gelatin surface is simulated successfully, which is in qualitative agreement with the experiment by Kodama and Takayama [Ultrasound in Med. & Biol. **24**, 723–738 (1998)]. It is also shown that the impulsive pressure at the bone surface caused by the bubble collapse is higher than that at the gelatin surface because the bubble collapse is accelerated by the high-pressure field generated by the reflection of the incident shock wave at the bone surface.

3:40

4pBB7. Acoustic characterization of echogenic liposomes: Attenuation and quantitative backscatter. Jonathan A. Kopechek (Dept. of Biomedical Eng., Univ. of Cincinnati, 231 Albert Sabin Way, MSB 6163, Cincinnati, OH 45267, kopechja@uc.edu), Tyrone M. Porter (Boston Univ., Boston, MA), Constantin-C. Coussios (Univ. of Oxford, Oxford, UK), Stephen R. Perrin, Jr. (Univ. of Cincinnati, Cincinnati, OH), Shaoling Huang, David D. McPherson (Univ. of Texas Health Sci. Ctr., Houston, TX), and Christy K. Holland (Univ. of Cincinnati, Cincinnati, OH)

Echogenic liposomes (ELIPs) are being developed for use as ultrasonic contrast agents and as drug carriers for ultrasound-targeted drug delivery. Physical and acoustical characterization of ELIPs is necessary in order to determine the optimum parameters for diagnostic and therapeutic applications. In this study, ELIP samples at concentrations of 10, 20, and 50 μg/ml were exposed to ultrasound in pulse-echo mode at center frequencies of 2.25, 3.5, 7.5, 10, 15, and 30 MHz. The received echoes were analyzed to determine the attenuation and backscatter coefficients and the results were compared to a theoretical computational model. The sample chamber contained two 50-μm tungsten wires as reference scatterers. The echoes from the wires were acquired before and after addition of ELIP to determine the attenuation coefficient. The backscatter coefficient was determined by averaging the square of the RF amplitude between the wires and accounting for transducer parameters. Each transducer was calibrated and characterized in deionized water using PVDF hydrophones. The peak attenuation coefficient occurred at 7.5 MHz while the backscatter coefficient increased with frequency. This was in agreement with the computational model. These results provide important information for determining the optimum acoustical parameters for ELIP exposure in diagnostic and therapeutic applications. [Work supported by NIH 2RO1 HL059586-04A2 and ASA Hunt Postdoctoral Research Fellowship.]

4:00

4pBB8. Delivery of targeted echogenic liposomes in an *ex vivo* mouse aorta model. Kathryn E. Hitchcock, Jonathan T. Sutton (Dept. of Biomed. Eng., Univ. of Cincinnati, 231 Albert Sabin Way, MSB 6163, Cincinnati, OH 45267, hitchcke@email.uc.edu), Danielle N. Caudell, Gail J. Pyne-Geithman, D. Phil. (Univ. of Cincinnati, Cincinnati, OH), Melvin E. Klegerman, Shaoling L. Huang, Deborah Vela, David D. McPherson (Univ. of Texas Health Sci. Ctr., Houston, TX), and Christy K. Holland (Univ. of Cincinnati, Cincinnati, OH)

Optimal ultrasound parameters to enhance delivery of therapeutic-loaded echogenic immunoliposomes (ELIP) into the arterial wall are being developed for the treatment of atherosclerosis. The aim of this work was to determine whether anti-ICAM-targeted, rhodamine-labeled ELIP (Rh-ELIP) would adhere to and penetrate the vascular endothelium in atheromatous murine arterial segments with intravascular flow treated with 1-MHz continuous wave ultrasound (CW US). A broadband focused hydrophone, confocally aligned with the artery and 1-MHz transducer field was used as a passive cavitation detector (PCD). Arteries were insonified with 1-MHz CW US (0.49 MPa peak-to-peak pressure), and the PCD was used to verify the duration of the resulting stable cavitation. Perivascular saline was collected and analyzed spectrofluorometrically for the presence of Rh-ELIP. Arteries were prepared for histological analysis by a pathologist blinded to the exposure conditions. Arteries exposed to Rh-labeled ELIP and 1-MHz US exhibited greater adherence of Rh-ELIP to the vascular endothelium and greater passage of Rh-ELIP across the vessel wall. No damage was detected in any of the arteries on histology. These studies will aid in the development of a strategy for improving atheroma treatment without causing ultrasound-related tissue damage.

4:15

4pBB9. 120 kilohertz ultrasound-enhanced thrombolysis in a porcine intracerebral hemorrhage model. Azzdine Y. Ammi (Dept. of Biomedical Eng., Colleges of Medicine and Eng., Univ. of Cincinnati, 231 Albert Sabin Way, Cincinnati, OH 45267, ammia@ohsu.edu), Saurabh Datta (Siemens, CA), Stephen R. Perrin, Jr, Shauna L. Beiler, Christian R. Beiler, Kenneth R. Wagner, and Christy K. Holland (Univ. of Cincinnati, Cincinnati, OH 45267)

Ultrasound acts synergistically with thrombolytic agents, such as recombinant tissue plasminogen activator (rt-PA), to accelerate thrombolysis. The aim of the study was to demonstrate the efficacy of 120-kHz ultrasound-enhanced rt-PA thrombolysis in a porcine hemorrhagic stroke model *in vivo*. Clots were formed by infusing 3 ml of autologous blood into the frontal white matter of 30 mixed-bred Yorkshire pigs (20.5–3.1 kg) and incubated for 3 h. For these nonsurvival studies, six pigs received rt-PA alone (0.3 cc of 0.107 mg/ml), six received ultrasound alone, six received rt-PA plus ultrasound, six were sham-exposed (saline only), and six were controls (no ultrasound or rt-PA treatment). The clots receiving ultrasound treatment were insonified with a peak-to-peak pressure of 0.48 MPa *in situ* (80% duty cycle, and PRF of 1.7 kHz) for 30 min. Clots treated with rt-PA alone exhibited a volume loss of 55.0% and clots treated with rt-PA and 120-kHz ultrasound had a significantly higher volume loss of 75.2% and a higher penetration of rt-PA. Thus, 120-kHz pulsed ultrasound enhancement of thrombolysis has been demonstrated both *in vitro* and in an *in vivo* porcine hemorrhagic stroke model.

4:30

4pBB10. Modelling of oscillations of a microbubble in an elastic vessel. Sergey Martynov, Eleanor Stride, and Nader Saffari (Univ. College London, Torrington Pl. London WC1E 7JE, UK, s.martynov@ucl.ac.uk)

Encapsulated microbubbles have been extensively investigated as contrast agents for diagnostic ultrasound imaging and more recently for therapeutic applications such as drug delivery. However, theoretical models for microbubble dynamics exist either for encapsulated bubbles in an infinite volume of liquid, or for unencapsulated bubbles in a confined volume. In the present study, a finite-element method is applied to quantify the effects of both encapsulation and confinement in a blood vessel upon a microbubble's response to ultrasound. The effect of encapsulation is examined for polymeric and surfactant coatings. Elastic deformations of the vessel wall are described using a lumped-parameter model, treating the wall as a thin membrane. It will be shown that even at low acoustic pressures (10 kPa), the bubble oscillations can be significantly modified as a result of confinement. In particular, the frequency spectrum of the oscillations of a confined bubble is characterized by two modes. For relatively soft vessels, a high-frequency mode dominates, with the eigenfrequency increasing with the vessel stiffness. The eigenfrequency of the low-frequency mode decreases with the vessel length. The results will be discussed in the context of diagnostic and therapeutic applications.

4:45

4pBB11. Electrocardiogram-gated imaging of a mouse heart using a high-frequency annular array. Jeffrey A. Ketterling, Jonathan Mamou (Riverside Res. Inst., Lizzi Ctr. for Biomedical Eng., 156 William St., New York, NY 10038), and Orlando Aristizábal (Skirball Inst. of Biomolecular Medicine and New York Univ. School of Medicine, New York, NY 10016)

Gated imaging is a technique in which M-mode data, referenced to a feature of an ECG signal, are acquired at a series of lateral positions and then reassembled into B-mode images to achieve high effective frame rates. It is only effective for imaging objects with periodic motion. Annular arrays and synthetic focusing allow for an improved depth of field (DOF) and resolution versus single-element transducers. Here, a five-element, 35-MHz annular array was utilized in combination with ECG and respiratory gating to image adult mouse hearts. An experimental system was assembled to permit appropriate triggering conditions and to collect all 25 transmit-to-receive echo data sets from the annular array at a series of lateral positions. The system was initially tested with a hexagon-shaped target that rotated at 10 revs/s and generated a trigger each rotation. After system testing, data were acquired from an adult mouse heart using a trigger in phase with the R-wave of the ECG signal. The synthetically-focused B-mode images showed several cardiac cycles over a 1200 ms duration captured at 100 fps over a 1-cm depth and a 6-mm width. Data from the mouse heart were acquired with conventional monocycle excitation and also with coded excitation.

5:00

4pBB12. High-frequency imaging with targeted ultrasound contrast agents under vascular flow conditions. Pavlos Anastasiadis (Dept. of Mech. Eng., Univ. of Hawaii at Manoa, 2540 Dole St., Honolulu, HI 96822, pavlos@hawaii.edu) and John S. Allen (Univ. of Hawaii at Manoa, Honolulu, HI 96822)

Targeted ultrasound contrast agents (UCA) were conjugated with biotinylated ligands, which allow them to adhere to specific diseased sites of interest. The overall adhesion of contrast agents is influenced by both the flow mediated gene expression and the local hydrodynamic forces acting on the contrast agents' trajectories. The ability to bind under pulsatile flow conditions and scatter sound at high frequencies is important for potential applications involving intravascular ultrasound. Human aortic endothelial cells (HAECs) in a flow chamber were exposed to a steady shear pretreatment over 12 h using a peristaltic pump system and subsequent pulsatile waveforms. To stimulate an inflammatory reaction in the HAECs, cells were subsequently exposed to rhTNF945; (Roche Pharmaceuticals). Different flow rates were applied, and the binding efficacy of targeted UCA was evaluated. Real-time transendothelial electrical impedance measurements and simultaneous acoustic measurements were performed on the same specimen with a scanning acoustic microscope. Applications related to targeting plaque in blood vessels are discussed. [This work was supported by the National Institutes of Health Grants NIH 2 P20 RR016453-05A1 and NIH 2 G12 RR0030161-21.]

Session 4pNSa

Noise, Architectural Acoustics, and ASA Committee on Standards: Soundscape Techniques and Applications—Community and Urban Environments

Brigitte Schulte-Fortkamp, Cochair
Inst. of Fluid Mechanics and Engineering, Technical Univ. Berlin, 10587 Berlin, Germany

Bennett M. Brooks, Cochair
Brooks Acoustics Corporation, 30 Lafayette Sq., Suite 103, Vernon, CT 06066

Chair's Introduction—1:00

Invited Papers

1:05

4pNSa1. Soundscape approach in progress—Report on the current stage. Brigitte Schulte-Fortkamp (TU-Berlin, Inst. of Fluid Mech. and Eng. Acoust., Einsteinufer 25, D-10587 Berlin, Germany)

Soundscapes have become an important issue in environmental acoustics. Following Soundscape workshops in Vancouver, Salt Lake City, and Miami, a Soundscape Symposium took place in Berlin, Germany, and brought together expertise from all over the world to work on concepts, approaches, analyses, applications, as well as source- and pattern recognition with respect to Soundscapes. Therefore, acousticians, architects, city planners, engineers, psychologists, sociologists, and the people concerned were invited. The aim of the workshop was to define Soundscapes for future work. Moreover, concepts, approaches, analyses, and applications were related to the categorization of Soundscapes as urban, cultural, and wilderness, also focussing on source and pattern recognition. This paper will report on the results to outline the next steps towards standardization.

1:25

4pNSa2. Integrating soundscape analysis into the National Environmental Policy Act process: A case study. George Luz (Luz Social and Environ. Assoc., 4910 Crowson Ave, Baltimore, MD 21212, Luz-Assoc.@msn.com)

A challenge for the proponents of soundscape analysis within the U.S. is how to integrate subjective observations with the quantitative requirements of the National Environmental Policy Act (NEPA), which, historically, emphasize physical measurements of all pollutants, including sound pollution. This paper describes a case study in which an attempt is made to integrate the subjective impressions of a soundscape analysis with objective measurements of equivalent sound level (LEQ). The occasion for this effort was a Base Realignment and Closure (BRAC) action in which the proposed reuse for a former military installation is park and recreational use. The quantitative framework for this study was a graphic used by the U.S. Army Environmental Hygiene Agency (USAEHA) during the 1980s in which the 24 h pattern of outdoor noise exposure is displayed as the minimum, average, and maximum values of 10 min LEQ over the course of a week of measurements. Although the intent of the USAEHA approach had been to determine conformance of military residential areas to Army guidelines, their method for displaying measurement data proved to be amenable to soundscape analysis as well.

1:45

4pNSa3. Soundscaping the aesthetic—Beyond measurement and assessment. Alexander U. Case (Sound Recording Technol., Univ. of Massachusetts Lowell, 35 Wilder St., Lowell, MA 01854, Alex_Case@uml.edu)

Soundscape in the context of community noise alone risks missing the important subjective reactions people have to the aesthetic value of any sound environment. The music recording and film sound industries have been recording sounds for art and entertainment, without regard for objective assessment, for more than a hundred years. Soundscape design has the opportunity to borrow from the time-tested recording, signal processing, and playback techniques developed by recording engineers where appropriate, learning from their search for emotional, intellectual, and uniquely human reactions. An introduction to nonobjective recording techniques is presented, with emphasis on some of the more counter-intuitive parts of the recording craft.

Contributed Paper

2:05

4pNSa4. A study on the modified urban soundscape of a city due to introduction of elevated structures. Kalaiselvi Ramasamy and Ramachandraiah Alur (Dept. of Civil Eng., IIT-Madras, India)

The urban soundscape of cities in a developing country like India are slightly varied by virtue of the fact that the composition of the traffic is heterogeneous accompanied by variance in road geometry and varying density of the buildings on the either side of the road and other community noise sources. Urban planning plays a vital role in organizing a city's traffic flow. To avoid congestion of traffic streams introduction of flyovers in the traffic flow happens to be a common feature in many urban environments. Through introduction of flyovers an increase or decrease in environmental noise characteristics occurs. In this paper a noise mapping study has been attempted along with field measurements of L10, L50, L90, and Leq to understand the soundscape of the city due to such types of modified topography. It also describes how the local characteristics of the city and changed topography alters the soundscape of the city. The theoretical computation of noise levels has been carried out using the sound plan software. It is observed that a reasonable reduction of Leq occurs in the immediate vicinity of noise sensitive areas of significant buildings such as hospitals and educational campuses.

THURSDAY AFTERNOON, 21 MAY 2009 EXECUTIVE SALON II/III, 2:30 TO 5:10 P.M.

Session 4pNSb

Noise, Architectural Acoustics, and ASA Committee on Standards: Soundscape Techniques and Applications—Wilderness and Park Soundscapes

Nancy S. Timmerman, Cochair
Timmerman Consultant in Acoustics, 25 Upton St., Boston, MA 02118-1609

Paul D. Schomer, Cochair
Schomer & Associates Inc., 2117 Robert Dr., Champaign, IL 61821

Chair's Introduction—2:30

Invited Papers

2:35

4pNSb1. Overview of on-going Federal Aviation Administration and National Park Service collaborative research efforts in support of the National Parks Air Tour Management Act. Cynthia Lee (U.S. Dept. of Transportation, Res. and Innovative Technol. Admin., Volpe Ctr., 55 Broadway, RVT-41, Cambridge, MA 02142, Cynthia.Lee@dot.gov)

In support of the National Parks Air Tour Management Act of 2000 (NPATMA), the FAA and NPS are developing air tour management plans (ATMPs) for approximately 100 national parks. ATMP objectives are to develop acceptable and effective measures to mitigate or prevent significant adverse impacts, if any, of commercial air tour operations upon the natural and cultural resources of and visitor experiences in national parks and abutting tribal lands. In accordance with NPATMA, any methodology adopted by a federal agency to assess air tour noise under this Act shall be based on reasonable scientific methods. Both agencies acknowledge that additional research is needed. This paper presents an overview of Volpe Center contributions to the collaborative effort by the FAA and NPS to develop improved methods to (1) quantify park soundscapes (natural and non-natural); (2) enhance computer modeling capabilities through aircraft source noise database expansion and advanced research into the factors which affect aviation noise propagation for complicated environments, such as national parks; and (3) assess the effects of aircraft overflights on park visitor experience, including the metrics used in these assessments. Improved methods developed in these efforts will also be used to support other agency projects related to aviation noise.

2:55

4pNSb2. Visitor perception of park soundscapes: An approach and research plan. Paul Schomer (Schomer and Assoc., Inc., 2117 Robert Dr., Champaign, IL 61821, schomer@SchomerAndAssoc..com) and G. Randy Stanley (Natl. Park Service, Fort Collins, CO 80525)

In the United States, much of the research on the soundscape in national parks and wilderness areas has centered on so-called dose-response relations with the aim of stating: How much manmade noise is unacceptable? Based on their experience in urban areas, this approach has been advocated primarily by the aviation noise community. But there is another school of thought. Visitors to national parks are guests and patrons of what the park has to offer. Just as juries of listeners are used to judge the sound quality of automobiles and home-appliances, juries of listeners could also be used to judge the sound quality in national parks. The current plan is to develop outdoor sound quality measurement and prediction methods and standards by which the U.S. National Park Service can accomplish their assessments of the acoustic ambient. This paper discusses the plan for this measurement protocol development and indicates some standards to be developed.

3:15

4pNSb3. "The Grand Canyon" versus "Soundscape of Nowhere (continued)". Dickson J. Hingson (Sierra Club—Natl. Parks and Monuments Committee)

More than 21 years have elapsed since the National Parks Overflights Act mandated the prompt "substantial restoration" of the natural quiet of the aircraft-noise imperiled soundscape of the Grand Canyon National Park. However, as of 2008, long-anticipated, critical compliance benchmarks have still not been timely met in the Park. The past two Administrations have not conformed to specifications/standards/deadlines set or appropriate to the NPS under its legal mandates. However, every battle has its turning point. *Will 2009 be the turning point to a quiet Canyon?* Success will require immediate NPS application of long-established restoration standards (based on "audibility"), Park zoning considerations, and buttressing with emerging supplemental noise indicators, which trigger loudness and temporal impulsiveness mitigations. Effectiveness of imminently anticipated management actions in the form of a soon forthcoming 2009 environmental impact statement and stepped up political oversight will be examined. These will pit restoration of the authentic Grand Canyon wilderness soundscape against the current, unsavory option: "the Soundscape of Nowhere." The protracted Grand Canyon imbroglio illuminates similarly unmet, pressing restoration needs, along with the need for increased executive /congressional oversight, re low-altitude air tour noise unacceptably continuing at similarly impacted, famed national parks, which otherwise remain subject to long-term, aviation noise impairment.

3:35

4pNSb4. A case for the importance of context for soundscape research in parks and protected areas. G. (Randy) Stanley (Natural Sounds Prog., U.S. Natl. Park Serv., 1201 Oakridge Dr., Ste. 100, Ft. Collins, CO 80525, Randy_Stanley@nps.gov) and Paul Schomer (Schomer and Assoc., Inc., Champaign, IL 61821)

Over the past decade or so, there have been a number of research studies, summarized herein, that indicate context is a potentially important intervening variable in assessing the soundscape in a park or wilderness area. Context is primarily provided by the corresponding visual landscape but may, in testing situations, be provided in other ways such as a verbal or written description. Context has been shown to be important in sound quality testing of automobile and product sounds, but its importance in assessing environmental sounds is less well documented. This paper discusses these various issues and suggests possible courses of action.

Contributed Papers

3:55

4pNSb5. The acoustical status of U. S. national parks. Kurt Fristrup (NPS Natural Sounds Program, 1201 Oakridge Dr., Ste. 100, Fort Collins, CO, 80525)

Acoustical monitoring data have been collected for more than 40 national park units, spanning a wide range of regions, resources, and park purposes and values. Comparative analyses of these data reveal the extent of noise pollution in national parks. Transportation noise is audible in many wilderness areas more than 30% of the daylight hours, with peak hourly levels approaching 70%. Peak levels from noise events can be more than 50 dB above the natural background under air tour routes, and between 10–20 dB above background for high altitude aircraft. Hourly Leq values can be as much as 6 dB above natural levels. These data also reveal the degree to which various noise metrics covary in these natural settings. This information helps inform selection of a compact set of metrics that can assess several functional impacts of noise while minimizing redundancy.

4:10

4pNSb6. Baseline sound monitoring plan for Grant-Kohrs Ranch national historic site. Robert C. Maher (Elec. & Comput. Engr., Montana State Univ., 610 Cobleigh Hall, Bozeman, MT 59717-3780, rob.maher@montana.edu)

Grant-Kohrs Ranch National Historic Site (GRKO), located just north of Deer Lodge, Montana, is a working cattle ranch commemorating the heritage of cowboys and stock growers in the history of the American West during the 19th and 20th centuries. The U.S. National Park Service (NPS) maintains the site according to its charter as a working ranch, with all the sights, sounds, and sensations associated with ranching. The cultural soundscape of the working ranch is considered essential to visitor enjoyment and understanding. Several anticipated changes in the neighboring community of Deer Lodge may affect the visitor experience at GRKO, including anticipated expansion of the local airport, increasing interstate highway traffic, and proposals to establish a rifle shooting range nearby. Because GRKO currently has no data characterizing the natural and cultural sounds of the park, this project was commissioned to monitor and evaluate the natural, cultural, and community sounds that comprise the ambient acoustic environment of the historic site over the period of one calendar year. The baseline acoustical data are analyzed and documented in a format suitable for management purposes by the NPS (the sponsor of this study), and by the NPS Natural Sounds Program Office.

4:25—5:10 Panel Discussion

Session 4pPA

Physical Acoustics, Underwater Acoustics, and Engineering Acoustics: A Half-Century with the Parametric Array II

Kenneth G. Foote, Cochair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Murray S. Korman, Cochair
Physics Dept., U. S. Naval Academy, Annapolis, MD 21402

Contributed Papers

1:20

4pPA1. Evaluation of parametric array technology for acoustic landmine detection. Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402, korman@usna.edu), Antal A. Sarkady and Frederick R. Tolle (U.S. Naval Acad., Annapolis, MD 21402)

There has been interest in using the parametric array for obtaining a highly directional low frequency source in acoustic landmine detection [M. S. Korman, J. Acoust. Soc. Am. **122** (2007)]. Earlier experiments used a 0.5 m commercial parametric array made up of 70 elements located 1.5 m directly over the target. The array was driven by a 110–1100 Hz swept sine audio modulated 65 kHz tone. A VS 1.6 inert plastic antitank landmine was buried 2.5 cm deep in dry sifted masonry sand in a concrete box. The laser Doppler vibrometer to microphone rms response was sufficient to measure the "on target" to "off target" contrast ratios of 20 and 3 for peaks near 850 and 1050 Hz, respectively (upon signal averaging) but missed the largest peak near 150 Hz (where the SPL was 40 dB per 20 μPa) among others. Recent "forward looking" experiments aligned the array beam axis at 30° from grazing. Sweeping from 110–210 Hz (1600 points) at 16 s/sweep and signal averaging 20 sweeps produced a contrast ratio of 10 for resonance at 148 Hz. Considerable improvement in SPL is necessary in order to make this technology practical for low frequency applications. [Work supported by ARL.]

1:35

4pPA2. Spatial phase-inversion technique for parametric source with suppressed carrier. Tomoo Kamakura, Hideyuki Nomura, and Shinichi Sakai (Dept. of Electronis, Univ. of Electro-Commun.s, 1-5-1, Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan)

Two planar projectors with the identical rectangular apertures are placed side by side. Both the projectors are radiating bifrequency ultrasound beams of finite amplitude in the air. The frequencies are 40 and 42 kHz but the initial phases are different. Especially, two extreme cases are considered: one is conventional in-phase driving, and the other is phase inversion driving. Sound pressure profiles were measured along and across the sound beam axis for the primary waves and the difference frequency wave of 2 kHz. The second and third harmonic components of the difference frequency were measured as well. Obviously, the pressure levels of the primary waves were suppressed considerably near the beam axis due to phase cancellation when the driving signals were out-of-phase by 180 degrees. The beam pattern of the difference frequency was, however, almost the same as the case where the signals were in phase. Interestingly, the pressure levels of the harmonics were reduced more than ten decibels. The validity of experimental results has been supported by good agreement with the theoretical predictions based on the Khokhlov-Zabolotskaya-Kuznetsov model equation. [Work supported by JSPS.]

1:50

4pPA3. Infrasound-convection nonlinear interaction. Konstantin Naugolnykh (Earth System Res. Lab., NOAA/Univ. of Colorado/Zeltech, Boulder, CO 80305)

The temperature stratified atmospheric layer is unstable and convection flow can be developed in such an area. Convection happens because warm less dense air goes up while cooler air comes down. The presence of infrasound produces modulation of convection flow and the sound wave amplification. The process of infrasound-convection nonlinear interaction is considered in the present paper. The equations of a compressible fluid convection are derived, which is characterized by the modified Rayleigh number, in comparison to this number for the incompressible flow, and condition of infrasound amplification is obtained. The increment of infrasound amplification turned out to be proportional to the Rayleigh number.

2:05

4pPA4. Sonar material acoustic property measurements using a parametric array. Victor F. Humphrey (Inst. of Sound and Vib. Res., Univ. of Southampton, Southampton, SO17 1BJ, UK, vh@isvr.soton.ac.uk.), Stephen P. Robinson, Graham A. Beamiss, Gary Hayman (Natl. Physical Lab., Teddington, Middlesex, TW11 0LW, UK), John D. Smith (DSTL, Porton Down, Salisbury, Wiltshire, SP4 0JQ, UK), Michael J. Martin (QinetiQ Ltd., Farnborough, Hampshire GU14 0LX, UK), and Nicholas L. Carroll (QinetiQ Ltd., Park, Winfrith Newburgh, Dorchester, Dorset, DT2 8XJ UK)

The use of a parametric array as an acoustic source for measuring the acoustic properties of materials for use in sonar systems is considered. Techniques of measuring the wideband transmission and reflection properties of limited size test panels are described and the advantages of using a parametric array in terms of reducing edge diffraction errors are discussed. Results for a range of materials under ambient pressure over the range 10–200 kHz are presented to illustrate the potential of the technique. The implementation of the approach in a pressure vessel at the UK National Physical Laboratory is also described and illustrated with example results obtained over the frequency range 2–50 kHz for two test objects that have predictable behavior. The potential of the technique is also illustrated with experimental results for viscoelastic test panels for hydrostatic pressures up to 2.8 MPa.

2:20

4pPA5. Parametric array application for long range ocean sounding. Konstantin Naugolnykh (Earth System Res. Lab., NOAA/Univ. of Colorado /Zeltech, Boulder, CO 80305) and Igor Esipov (N. Andreev Acoust. Inst., 4 Schvernik St., Moscow, Russia)

The parametric array (PA) is a nonlinear transducer that generates narrow, sidelobe-free beams of low frequency sound, through the interaction of high frequency pump waves. PA was suggested by Peter J. Westervelt, winner of the Lord Rayleigh Medal, at the same time this devise invention was underway in the Soviet Union made by Zverev and Kalachev. Remarks of PA applications for long distance ocean sounding is presented in the present paper. Experimental test of a high-frequency PA was made on the Black Sea coast by Esipov *et al.*, and investigations of powerful PA characteristics was performed in a ocean by Andebura *et al.*, in 1990. Then large-scale ocean vortices sensing at a distance of 1000 km was realized by Esipov *et al.* [Acoust. Zh., **40**(1), 71–75 (1994)]. In both experiments, the transducer of R/V "Boris Konstantinov" was used with parameters: pump wave power

$W=20$ kW, pump wave frequency $f=3$ kHz, parametric signal frequency $F=230$–700 Hz. Obtained results indicate the efficiency of PA application for long distance ocean sounding. [Work supported by ISTC project 3770 and NATO Grant No. ESP.NR.NRCLG 982524.]

2:35

4pPA6. Observing Atlantic herring by parametric sonar. Olav Rune Godø (Inst. of Marine Res., PO Box 1870, N-5817 Bergen, Norway, olav.rune.godoe@imr.no), Kenneth G. Foote (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), Johnny Dybedal (Kongsberg Defence & Aerosp. AS, Stjoerdal, Norway), and Eirik Tenningen (Inst. of Marine Res., Bergen, Norway)

Atlantic herring (*Clupea harengus*) has been observed in situ by the Kongsberg TOPAS PS18 parametric sub-bottom profiling sonar, with nearly horizontal transmit and receive transducer arrays mounted flush on the hull of R/V G. O. Sars, at N71.4 E16.3. The primary frequencies are in the band 15–21 kHz, with nonlinearly generated difference frequencies in the band 0.5–6 kHz. The beamwidths are 6–10 deg depending on frequency and range, but with exceedingly low sidelobes. The observation of herring in schools and layers was accomplished with the vessel both at rest and sailing at its ordinary survey speed of 10 knots. The observations with the parametric sonar were confirmed by concurrent, synchronized observations with the Simrad EK60/38-kHz scientific echo sounder and by trawling with a pelagic net. The herring length varied from 28.0–36.5 cm. Present work suggests that parametric sonar will become a powerful new tool in marine ecosystem studies, enabling the numerical density of schooling and shoaling fish to be determined, and the size of swimbladder-bearing fish to be estimated by detection of swimbladder resonance. [Work partly supported by Norwegian Research Council Grant no. 184705.]

2:50

4pPA7. Range compensation function for echo integration in transducer near fields, with special reference to parametric sonar. Kenneth G. Foote (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543, kfoote@whoi.edu)

Range compensation functions are applied to echo signals so that the corresponding range-compensated echo intensity equals the volume backscattering coefficient to within a multiplicative constant. For echo integration in a transducer far field, the transmit and receive pressure fields each decrease inversely with range r, and the well-known range compensation function, applied to intensity, is $r^2\times10^{\alpha r/5}$, where α is the absorption coefficient. For echo integration in a transducer near field, the range compensation function is that of the far field modified by a multiplying factor. This factor is, approximately, the ratio of the square of the surface integral of the field intensity at r due to extrapolation from the far field, assuming inverse range dependence of the amplitude, to the product of the surface integrals of transmit and receive intensities at r. This general range compensation function is developed for the special case of echo integration with a parametric sonar in which scatterers are ensonified in a region where the nonlinearly generated difference-frequency field is growing and echoes are received in the far field of a collocated, linearly operating transducer. It is evaluated numerically for the Kongsberg TOPAS PS18 parametric subbottom profiling sonar, with primary-frequency band 15–21 kHz and difference-frequency band 0.5–6 kHz.

3:05

4pPA8. Standard-target calibration of a parametric sonar over the difference-frequency band, 1–6 kilohertz. Kenneth G. Foote (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543, kfoote@whoi.edu), Johnny Dybedal (Kongsberg Defence & Aerosp. AS, N-7501 Stjoerdal, Norway), and Eirik Tenningen (Inst. of Marine Res., N-5817 Bergen, Norway)

The Kongsberg TOPAS PS18 parametric sub-bottom profiling sonar operates over the frequency band 15–21 kHz, with nonlinearly generated difference-frequency radiation in the band 0.5–6 kHz. The TOPAS transducer mounted on R/V G. O. Sars is flush with the hull in the near-horizontal plane. The sonar has been calibrated by the standard-target method using a 280-mm diam sphere of aluminum alloy 6082 T6 [K. G. Foote *et al.*, J. Acoust. Soc. Am., **121**, 1482–1490 (2007)]. The target was suspended beneath the vessel at each of three ranges, successively 100, 200, and 300 m. Because of conditions in Soerfolla fjord on Dec. 10, 2008, the target sphere was moving slowly relative to the vessel. Its instantaneous position was determined by geometrical considerations through synchronous observation with the Simrad EK60/38-kHz scientific echo sounder, with split-beam transducer mounted approximate to the TOPAS transducer. Data were collected for a number of parameter settings for each of three signal types: continuous wave, chirp, and Ricker pulse. Measurements are compared with predictions based on laboratory measurements of the frequency-dependent sensitivities of the parametric transmitter and conventional linear receiver, using a range-compensation function based on theoretical nearfield modeling. [Work partly supported by Norwegian Research Council Grant No. 184705.]

3:20—3:40 Break

Invited Papers

3:40

4pPA9. Resonant mode conversion in superfluid 4 helium and in solids. Steven L. Garrett (Graduate Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

One component of the success of the Westervelt analysis of the parametric array was that sound propagation in most gases and liquids exhibit very little dispersion. Therefore, the pump waves and their nonlinearly generated product propagate at very nearly the same speed. When an acoustic medium can support more than one sound mode with different propagation speeds, then a parametric array can be created by the nonlinear interaction of the two slower sound waves to produce the faster wave, if the slow wave propagation directions are not collinear. Bulk superfluid helium, at temperatures below 2.1 K, will support a slower (thermal) sound mode and a faster (mechanical) sound mode. An experiment performed in 1977 will be described that used a waveguide to create thermal (temperature-entropy) sound waves while controlling their angle of intersection. A mechanical (pressure-density) wave was produced and its amplitude determined the thermodynamic coupling constant between density changes and the Galilean invariant square difference in the normal and superfluid particle velocities that characterize the thermal sound wave amplitude. An earlier experiment will also be described that demonstrated the nonlinear conversion of two shear waves to a longitudinal wave in aluminum. [Work supported by the Office of Naval Research.]

4:00

4pPA10. Quantum acoustics, the second law and the end fire array. Seth Putterman (Dept. of Phys., UCLA, Los Angeles, CA 90095) and Paul H. Roberts (UCLA, Los Angeles, CA 90095)

The fundamental nonlinear interaction which leads to the scattering of waves and to the end fire array yields a Boltzmann equation for sound, a one fluid theory of superfluid helium [S. Putterman, P. H. Roberts, Physics Letters, **89A**, 444 (1982)] and a route to the

quantum theory of interacting phonons [M. Cabot, S. Putterman, Phys. Lett. **83A**, 91 (1981)]. Along with a review of these consequences of nonlinear classical acoustics the conundrum of the consistency of the second law of thermodynamics with the Hamiltonian nature of the end fire array will be exposed but not resolved [S. Putterman, P. H. Roberts, Phys. Rep., **168**, #4 (1988)].

Contributed Paper

4:20

4pPA11. Acoustic nonlinearity in fluorinert. Cristian Pantea, Dipen N. Sinha, Curtis F. Osterhoudt, and Paul C. Mombourquette (Los Alamos Natl. Lab., MPA-11, MS D429, Los Alamos, NM 87545, pantea@lanl.gov)

Fluorinert FC-43 nonlinearity was investigated using two approaches: (i) a finite amplitude method with harmonic production and (ii) a nonlinear frequency mixing in the fluid with consequent beam profile measurement of the difference frequency. The finite amplitude method provides information on the coefficient of nonlinearity, β, through the amplitudes of the fundamental and the second harmonic, at a certain transmitter-receiver distance. A calibrated hydrophone was used as a receiver in order to obtain direct pressure measurements of the acoustic waves in the fluid. The role of transmitter-receiver distance in β determination is investigated. In the second approach, a single transducer is used to provide two high-frequency beams. The collinear high-frequency beams mix nonlinearly in the fluid resulting in a difference frequency beam and higher order harmonics of the primaries. The difference frequency beam profile is investigated at lengths beyond the mixing distance. The experimental data are compared with the KZK theory.

Invited Paper

4:35

4pPA12. Flow through orifices caused by large-amplitude asymmetric sound waves: Theory and experiment. Peter J. Westervelt (Dept. of Phys., Brown Univ., Providence, RI 02912, abwpjw@cox.net)

When the displacement amplitude of an acoustic wave consisting of equal parts fundamental and second harmonic exceeds the diameter of a circular orifice, steady flow is generated whose direction is determined by the phase of the two wave components [P. J. Westervelt, Ph.D. thesis "The interaction of a finite amplitude acoustic wave with small obstacles and orifices," (MIT, 1951)]. Some applications to semipermeable biological membranes are discussed. [Work supported by the Office of Naval Research].

THURSDAY AFTERNOON, 21 MAY 2009 GRAND BALLROOM I, 1:00 TO 4:00 P.M.

Session 4pPP

Psychological and Physiological Acoustics: Physiology; Aging; Hearing Loss (Poster Session)

Kathryn H. Arehart, Chair
Dept. of Speech, Language and Hearing Science, Univ. of Colorado at Boulder, Boulder, CO 80309

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 even-numbered papers will be at their posters from 2:30 p.m. to 4:00 p.m.

4pPP1. Effect of stimulus onset delay on auditory cortex neural responses to voice pitch feedback perturbation. Roozbeh Behroozmand and Charles R. Larson (Speech Physiol. Lab., Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208)

It has previously been shown that the auditory neural responses to voice F0 feedback perturbation are suppressed during active vocalization compared to passive listening to the playback. However, a study in primates showed that the vocalization-induced suppression enhances auditory sensitivity to feedback perturbation. This evidence suggests that the cortical neural responses to self-vocalization reflect suppression to vocal onset and excitation in response to perturbations in voice pitch feedback. In this study, we investigated the effect of stimulus onset delay on cortical neural responses to voice F0 feedback perturbation. Event-related potentials (ERPs) were recorded in human subjects in response to simultaneous (0 ms) and delayed (500 ms) PSS (+200 cents) during active vocalization and passive listening conditions. Results showed that, for the delayed PSS, the P200 peak magnitude was larger during vocalization compared with passive listening. This finding suggests that vocalization enhances auditory cortex responsiveness to deviations in voice pitch feedback following the onset of vocalization. This enhancement might be due to changes in tuning properties of auditory neurons by the vocal motor system that helps to detect and correct for vocal errors during speech production.

4pPP2. Event-related potential correlates of online monitoring of auditory feedback during vocalization. Colin S. Hawco, Jeffery A. Jones, and Todd R. Ferretti (Dept of Psych., Wilfird Laurier Univ., 75 University Ave. West, Waterloo, ON N2L 3C5, Canada)

When speakers hear the fundamental frequency (F0) of their voice altered, they shift their F0 in the direction opposite to the perturbation. The neural mechanisms underlying this response are poorly understood. In the present study, event-related potentials (ERPs) were used to examine the neural mechanisms used to detect alterations in auditory feedback during an ongoing utterance. Participants vocalized for 3 s, and heard their auditory feedback shifted by 0, 25, 50, 100, or 200 cents for 100 ms midutterance. In two sessions, participants either vocalized at their habitual pitch, or matched a target pitch. A mismatch negativity (MMN) was observed, with the amplitude positively related to the size of the perturbations. No differences were found between sessions. The F0 compensation response was found to be smaller for 200 cent shifts than 100 cent shifts, and a positivity was observed in the ERPs for a 200 cent shift. This result suggests that a 200 cent shift may be perceived as externally (rather than internally) generated. The

presence of an MMN, and no earlier (N100) response suggests that the underlying sensory process used to identify and compensate for errors in midutterance may differ from feedback monitoring at utterance onset.

4pPP3. Ictal and interictal changes in central auditory processing. David B. Daly and David M. Daly (Box 210855, Dallas, TX 75211, dave.daly @stanfordalumni.org)

Disordered functioning manifest in seizures can also give rise to subtle, pervasive interictal changes. We have used sets of synthesized acoustic stimuli, concurrent EEG and AED blood-level monitoring to evaluate changes in a 24-year old female with focal seizures. During ictus patient becomes mute, she can hear but not comprehend, and may then be briefly amnestic. EEG revealed sharp/slow waves over left anterior temporal regions, with occasional bilateral discharges. Levels of two AED fluctuated 50%, with peaks 2 h apart. Performance on 4-min sets of auditory testing with ge-ye and be-de-ge fluctuated from well defined ($p<0.001$) near either peak AED, to near-chance levels; three-choice vowel sets were well defined throughout. Disruptions, which involved contralateral homologous areas as well as surrounding ipsilateral areas, are consistent with augmented inhibition.

4pPP4. Measures of wideband power reflectance in otosclerotic ears. Marcin Wróblewski (Graduate Ctr., City Univ. of New York and Dept. of Otolaryngol., NYU Langone Medical Ctr., 550 1st Ave., NBV 5E5, New York, NY 10016, marcin.wroblewski@nyumc.org), Arlene C. Neuman, Nancy Jiang, and Anil K. Lalwani (NYU Langone Medical Ctr., New York, NY 10016)

Wideband reflectance (WBR) is a new method of evaluating middle ear function. While several studies have reported WBR data obtained from persons with normal hearing and from children with otitis media, few data have been reported describing results typical of other types of middle ear pathology. The purpose of the present study was to collect data from a group of persons with clinically diagnosed otosclerosis and to compare the measures of WBR (including reactance, resistance, impedance magnitude, and transmittance) to measures obtained from persons with normal middle ear function. WBR data were collected from 17 preoperative and 18 postoperative otosclerotic ears as well as from 57 ears without middle ear pathology. Measures from four otosclerotic ears allowed for a direct comparison of pre- and postoperative middle ear function. There was a variety of responses from ears with otosclerosis. Preoperatively, while some otosclerotic ears showed a pattern reported in previous case studies (i.e., increased WBR below 1000 Hz indicating enlarged middle ear stiffness), others fell within the normal range. Postoperatively, some ears showed a pattern typical of an increased mass component of middle ear impedance (i.e., decreased WBR below 1000 Hz), while others fell within the normal range.

4pPP5. Comparison of measurements at ambient pressure on clinical immittance and wideband acoustic transfer function systems. Kim Schairer, Brooke Morrison, Ellyn Steininger, and Cynthia Fowler (Univ. of Wisconsin, 1975 Willow Dr., Rm. 373, Madison, WI 53706)

The purpose of the study was to compare acoustic admittance recorded using probe tones of 226, 678, and 1000 Hz on a clinical immittance system with admittance and reflectance recorded on a wideband acoustic transfer function (WATF) system. The WATF system uses a click probe, which yields measurements across a broad range of frequencies with one test, whereas multiple tests are required at individual frequencies on the clinical system. Thus, the WATF system has the potential to provide more information in a shorter amount of time. The hypothesis was that admittance would be comparable between the two systems, which would provide support for clinical use of the WATF system. Measurements were taken at ambient pressure (i.e., 0 Da Pa) in the ear canals of adults with normal hearing and middle-ear function. Because the clinical system reports only information at the tympanometric peak by default, the data at 0 Da Pa were manually recorded by entering the cursor mode. In general, the results at all three probe frequencies support the hypothesis, although the relationship was stronger for the two higher probe tone frequencies. The clinical implications and effect of analysis bandwidth on the WATF system will be discussed.

4pPP6. An electroacoustical analog for estimating sound pressure level at the tympanic membrane at high frequencies. Janice L. LoPresti (Knowles Electron.s LLC, 1151 Maplewood Dr. Itasca, IL 60143, janice.lopresti@knowles.com), Karrie Recker, and Tao Zhang (Starkey Lab., Inc., Eden Prairie, MN 55344)

For various applications, it is useful to know the sound pressure level (SPL) at the tympanic membrane (TM). However, measuring the SPL close to the TM is not clinically feasible, due to safety and discomfort concerns. Therefore, it is desirable to estimate the SPL at the TM using measurements away from the TM. This is challenging at high frequencies where the effect of canal geometry becomes significant as the wavelength of the acoustical source becomes comparable to the dimensions of the canal. In this study, an electroacoustical analog was used to estimate the SPL at the TM for ten participants. The analog is comprised of a transducer (i.e., sound source), occluded ear canal, and the middle ear [Pascal *et al.* (1998)]. The ear model was optimized for each individual using a 2-parameter fit using real ear measurements further away in the canal. The optimization did not require inputs that may be difficult to obtain such as the exact canal geometry. A simulated transfer function was created and applied to each measured real-ear response to generate the estimated response at the TM. The model was verified using the SPL measured at the TM.

4pPP7. Evidence for dynamic cochlear processing in otoacoustic emissions and behavior. 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)

Psychophysical research suggests that active cochlear processing may partially explain temporal effects observed in certain auditory masking tasks [Strickland (2001), (2004)]. To investigate this possibility, this study examined the relationship between subjects' psychophysical performance and the responses of their cochleas to the same stimulus waveforms. Stimulus-frequency otoacoustic emissions (SFOAEs) were recorded in the ear canal using a nonlinear procedure. The results showed that the nonlinear SFOAE to a brief tonal signal (4 kHz, 10 ms, 60 dB SPL) in a background noise (100–6000 Hz, 400 ms, 25 dB/Hz) increased in magnitude (with signal delay) over a similar time course to subjects' improvement psychophysically in detecting the same signal. Manipulation of the noise bandwidth revealed that the increases in SFOAE magnitude and the improvement in psychophysical detection both were caused primarily by off-frequency components of the noise: lowpass components for SFOAEs, and lowpass or highpass components for psychophysics. These findings are consistent with a change in the slope of the cochlear input-output function from highly compressive to less compressive over the duration of the background noise, a change that might be attributable to negative feedback from the efferent system. [Work supported by NIDCD.]

4pPP8. Parameters for the estimation of loudness from tone-burst otoacoustic emissions. Michael Epstein (Dept. of Speech-Lang. Path. and Aud., Ctr. for Commun. and DSP [ECE Dept.] Auditory Modeling and Processing Lab., Northeastern Univ., 1600 Massachusetts Ave., Boston, MA 02115, m.epstein@neu.edu) and Ikaro Silva (Northeastern Univ., Boston, MA 02115)

There is evidence that tone-burst otoacoustic emissions (TBOAEs) might be useful for estimating loudness, but appropriate analysis parameters for loudness estimation must first be examined. The purpose of the present work was to collect TBOAE measurements and loudness estimates across a wide range of levels in the same listeners. Both measures were recorded for 1- and 4-kHz stimuli and then analyzed using a wide range of parameters to determine which parameter set yielded lowest mean-square-error estimation of loudness with respect to a psychoacoustical, cross-modality-matching procedure and the inflected exponential (INEX) loudness model. The present results show strong agreement between 1-kHz loudness estimates derived from TBOAEs and loudness estimated using cross-modality matching, with TBOAE estimation accounting for a significant portion of the variance. Additionally, the results indicate that analysis parameters may vary within a reasonable range without compromising the results (i.e., the estimates ex-

hibit some parametric robustness). The lack of adequate parametric optimization for TBOAEs at 4 kHz suggests that measurements at this frequency are strongly contaminated by the ear-canal resonances, meaning that deriving loudness estimates from TBOAEs at this frequency is significantly more challenging than at 1 kHz. [Work supported by Capita Foundation.]

4pPP9. The sensitivity of hearing by the resonance in *in vivo* human ear canal. Wei-De Cheng (Taiouan Interdisciplinary Otolaryngol. Lab., Chang Gung Univ.), Jen-Fang Yu (Chang Gung Univ.), Kuo-Wei Huang, Shang-Peng Chang (Chang Gung Univ.), and Chin-Kuo Chen (Chang Gung Memorial Hosp.)

This study is to discuss the sensitivity of human hearing sounds at different depths of ear canal, and to construct the measurement scales which can be the reference for the sound measurement and research in the future. Eighteen subjects aged from 20 to 30 years old with normal hearing and middle ears were studied. The pure tone audiometer and impedance audiometer were utilized to exam the frequency threshold of subjects. The real ear measurement was also utilized. The intensities of stimuli were 40, 50, 60, 70, and 80 dB SPL. The measured depths to the tympanic membrane were 0.5, 1.0, 1.5, and 2.0 cm, respectively. The gain of ear canal by different frequencies at 500, 1000, 2000, 4000 Hz was measured. Based on the results, there was moderate negative correlation between the resonance of ear canal and hearing. The larger the gain of ear canal was, the better the hearing of the subject would be. Consequently, the sensitivity of hearing for normal people would be affected by the resonance of ear canal. The resonance of each subject would be changed similarly at different measurement depths.

4pPP10. The effects of quinine on frequency selectivity, temporal resolution, and speech recognition in quiet and noise. Erica J. Williams and Sid P. Bacon (Dept. of Speech and Hearing Sci., Arizona St. Univ., PO Box 870102, Tempe, AZ 85287-0102, ejw@asu.edu)

Quinine causes a temporary disruption of outer hair cell (OHC) function, and thus can be used to examine the role of OHCs on auditory perception. In the present study, frequency selectivity, temporal resolution, and speech recognition were measured before, during, and after a quinine-induced hearing loss. Normal-hearing listeners ingested 5.76–11.43 mg/kg body weight of quinine, resulting in 5–15 dB of hearing loss. Frequency selectivity was estimated by comparing the level of a noise masker needed to mask a fixed-level, 2-kHz signal when the masker contained a spectral gap at 2 kHz and when it contained no gap. Similarly, temporal resolution was estimated by comparing the masker level needed to mask the 2-kHz signal when the noise masker contained a temporal gap and when it did not. Speech recognition thresholds were measured in quiet and in the presence of a masker (speech-shaped noise or time-reversed speech) fixed at 70-dB SPL. Signal level was varied adaptively to estimate 50% correct recognition. Quinine resulted in reduced frequency selectivity and reduced temporal resolution. Quinine also elevated speech recognition thresholds in quiet (by 7 dB on average), but the thresholds in the presence of the maskers were unaffected. [Work supported by NIDCD and AAA.]

4pPP11. Behavioral responses to harmonic complex tones with missing fundamental frequencies by chinchillas in the presence of low-pass masking noise. William P. Shofner (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405, wshofner @indiana.edu)

The pitch associated with the missing fundamental (F0) is one of the principal psychological attributes of human pitch perception. Behavioral responses to missing F0 stimuli were measured in chinchillas using operant conditioning and stimulus generalization. Animals were trained to discriminate between harmonic complex tones having a 500-Hz F0 and a 125-Hz F0. When animals were tested with tone complexes having the same F0s, but where the F0s were missing, responses were similar to those obtained when the F0s were present, suggesting that missing F0 sounds were perceptually equivalent to F0 present sounds. In the presence of low-pass masking noise, responses to F0 present and missing F0 stimuli were similar, suggesting that the percept was not due to the reinsertion of the F0 through cochlear nonlinearities. When the F0s of test stimuli were systematically varied, gradients in behavioral responses were observed, suggesting the existence of a psychological dimension related to F0. When the F0 and spectrum were varied independently, responses were related to the F0 rather than to spectral differences among test stimuli. The results indicate that chinchillas possess a pitchlike perception of the missing F0 that is unlikely to arise from cochlear distortion products. [Work supported by NIDCD R01 DC005596.]

4pPP12. Temporal modulation transfer functions in listeners with real and simulated hearing loss. Joseph G. Desloge, Charlotte M. Reed, Louis D. Braida, Lorraine A. Delhorne, and Zachary D. Perez (Res. Lab. of Electron., Massachusetts Inst. of Technol., 77 Massachusetts Ave., Cambridge, MA 02139, jdesloge@mit.edu)

Temporal modulation transfer functions (TMTFs) were obtained in nine listeners with moderate to profound sensorineural hearing loss (age range of 21 to 69 years). The hearing loss of each impaired listener was simulated in a group of three age-matched normal-hearing listeners through a combination of spectrally-shaped masking noise and multi-band expansion. TMTFs in both groups of listeners were measured in a broadband noise carrier as a function of modulation rate in the range of 2 to 1024 Hz using a 3I, 2AFC procedure. The presentation level for the unmodulated broadband noise (whose duration was 500 ms) was set to be the maximum of either 70 dB SPL or the level such that noise was 10 dB above the lowest hearing threshold. The listeners with simulated hearing loss thus received signals at the same SPL and SL as their hearing-impaired counterparts. The shape of the TMTF curves (defined as the measured threshold of modulation in dB as a function of frequency of modulation) and the interpolated DC values of the function (using an exponential fitting procedure) were generally similar for listeners with real and simulated hearing loss. [Work supported by NIH-NIDCD R01 DC00117.]

4pPP13. Forward-masking functions in listeners with real and simulated hearing loss. Louis D. Braida, Joseph G. Desloge, Charlotte M. Reed, Lorraine A. Delhorne, and Zachary D. Perez (Res. Lab. of Electron., Massachusetts Inst. of Technol., 77 Massachusetts Ave., Cambridge, MA 02139, ldbraida@mit.edu)

Forward-masking functions were obtained in eight listeners with moderate to severe sensorineural hearing loss (aged 21–69 years). The hearing loss of each impaired listener was simulated in a group of three age-matched normal-hearing listeners through a combination of spectrally shaped masking noise and multiband expansion. Forward masking was measured in both groups of listeners for probe signals at frequencies of 500, 1000, 2000, and 4000 Hz using an on-frequency masker and two off-frequency maskers (0.55 and 1.15 times the signal frequency) under a 3I, 2AFC procedure. The probe signal was presented at 5 dB SL; signals and maskers were gated with 5-m on/off times with a steady-state duration of 0 ms for probe signals and 100 ms for maskers; and values of masker-offset time to signal-onset time were in the range of 0 to 100 ms. Findings will be described in terms of the slopes of the masking functions for each combination of probe and masker frequency and the ratio of the slopes obtained with on- relative to off-frequency maskers. Preliminary results suggest that the slope ratios observed for the hearing-impaired listeners were generally well-reproduced in normal-hearing listeners with simulated hearing loss. [Work supported by NIH-NIDCD R01 DC00117.]

4pPP14. A comparison of gap-detection thresholds through audition and touch in listeners with real and simulated hearing impairment. Charlotte M. Reed, Joseph G. Desloge, Zachary D. Perez, Lorraine A. Delhorne, and Louis D. Braida (Res. Lab. of Electron., Massachusetts Inst. of Technol., 77 Massachusetts Ave., Cambridge, MA 02139, cmreed@mit.edu)

Gap-detection thresholds were measured in listeners with real and simulated hearing loss under conditions of auditory or tactile presentation. The audiometric thresholds of each of ten listeners with sensorineural hearing loss (21–69 years of age) were simulated in groups of age-matched normal-hearing listeners through a combination of spectrally shaped masking noise and multiband expansion. The leading and trailing markers for the gap-detection task were 250- and 400-Hz sinusoids with a nominal duration of 100 ms. Gap-detection thresholds for a nominal baseline gap of 0 ms were measured for four different combinations of leading and trailing markers (250–250, 250–400, 400–250, and 400–400 Hz) using a 3I, 2AFC procedure. Auditory measurements were obtained for monaural presentation over headphones using a marker level set to be equal to the maximum of 70 dB SPL or 10 dB SL. Tactile measurements were obtained using sinusoids

presented to the left middle finger through an Alpha-M AV-6 vibrator at a level of 25 dB SL. Results are discussed in terms of (a) spectral-disparity effects of leading and trailing markers; (b) relation of thresholds for auditory versus tactile presentation; and (c) comparisons of results from listeners with real and simulated hearing loss. [Work supported by NIH-NIDCD R01 DC00117.]

4pPP15. Effects of several variables on temporal-order identification in young and elderly listeners. Daniel Fogerty, Diane Kewley-Port, and Larry Humes (Speech and Hearing Sci., Indiana Univ., 200 S. Jordan, Bloomington, IN 47405, dfogerty@indiana.edu)

Four measures of temporal-order identification were completed by young ($N=35$, 18–31 years) and elderly ($N=151$, 60–88 years) listeners under various stimulus conditions. Experiments spaced over several months used forced-choice constant-stimuli methods to determine the smallest stimulus onset asynchrony (SOA) between brief (40 or 70 ms) vowels that enabled identification of the stimulus sequence. Vowels in four words (pit, pet, pot, put) served as stimuli. The four measures of temporal-order identification were: (1) monaural two-item sequences; (2) monaural four-item sequences; (3) dichotic two-item vowel-identification sequences; and (4) dichotic two-item ear-identification sequences. All listeners identified the vowels in isolation with better than 90% accuracy. Results indicated that elderly listeners performed significantly poorer on monaural and dichotic temporal-order identification tasks than young listeners, although a large overlap in group distributions was observed. For both groups, the two-item dichotic task was significantly harder than two-item monaural. Increasing the attentional demands of the monaural task by randomizing the stimulus ear did not explain this difference. Using shorter duration stimuli did not alter performance in the monaural task but did improve performance in the dichotic task. Significant learning occurred for elderly listeners but not enough to eliminate the age-group differences. [Work supported, in part, by NIA.]

4pPP16. Level-dependent changes in detection of a silent gap in fluctuating noise carriers. Amy R. Horwitz, Jayne B. Ahlstrom, and Judy R. Dubno (Dept. of Otolaryngol.-Head and Neck Surgery, Medical Univ. of South Carolina, 135 Rutledge Ave., MSC 550, Charleston, SC 29425, horwitar@musc.edu)

Changes in detection of a silent gap in a noise carrier with increasing carrier bandwidth and level may reveal effects of basilar-membrane nonlinearities on envelope fluctuations. Gap-detection thresholds are higher for narrowband than broadband noise carriers due to increased confusion between the imposed gap and inherent fluctuations of the narrowband noise. For narrowband more than broadband carriers, the "effective" magnitude of envelope fluctuations may be reduced by compressive effects of the active cochlear mechanism. To test these assumptions, detection of a gap was measured in 50-Hz-wide and 1000-Hz-wide noise carriers as a function of carrier level. Younger and older adults with normal and impaired hearing listened to carriers presented in quiet and in a low-fluctuation broadband masker. As carrier level increased, masker level also increased, maintaining a fixed difference between carrier and masker levels to minimize sensation-level effects on gap detection. For younger adults, gap detection measured in quiet improved as expected with increasing carrier level. Further, gap detection measured in the masker improved as carrier level increased for the 50-Hz carrier, but declined for the 1000-Hz carrier. Discussion will focus on effects of carrier level, bandwidth, subjects' age and thresholds, and hypothesized relations to basilar-membrane nonlinearities. [Work supported by NIH/NIDCD.]

4pPP17. Minimum audible angles in children who use bilateral cochlear implants. Cynthia M. Zettler, Shelly Godar, and Ruth Y. Litovsky (Waisman Ctr., Univ. of Wisconsin-Madison, 1500 Highland Ave., Madison, WI 53705)

Over 5000 children worldwide have received bilateral cochlear implants (CI) so that the auditory skills that rely on inputs to both ears, such as sound localization, may be improved. In this study, localization acuity was measured with minimum audible angles (MAA), the smallest discriminable distance between two locations in the frontal azimuth plane. It is unclear whether children who use bilateral CIs can attain MAAs comparable to their acoustically hearing peers, and if so, the extent to which exposure to bilateral hearing is necessary. Children having 3–36 months of bilateral experience participated. Stimuli were spondaic words presented at an overall level of 60 dB ($\pm$ 4 dB rove). Every child with >5 yrs of auditory experience performed the task, whereas some of the children with <5 yrs of experience could not perform the task. In the former group, children with >2 yrs of bilateral experience were the best performers (MAA thresholds $<15°$), compared to children in the latter group who completed the task (MAA thresholds $>30°$). Results suggest that in children who receive bilateral CIs the two overall factors that impact performance are overall time in sound and amount of time with bilateral experience. [Work supported by NIH DC R01008365.]

4pPP18. The impacts of age and absolute threshold on binaural lateralization. Frederick J. Gallun, Anna Diedesch, and Erin Engelking (Natl. Ctr. for Rehabilitative Auditory Res., Portland VA Medical Ctr., Portland, OR, 97239, Frederick.Gallun@va.gov)

Listeners varying in audiometric thresholds (70-dB range) and in age (52-year range) lateralized complex tones relative to a diotic standard. Stimuli were composed of one, two, three, or six tones amplitude modulated at a rate of 75 Hz. Carrier frequencies were selected from a set of seven values (roughly log-spaced between 700 and 7000 Hz) and sensation levels were equated across listeners. Baseline performance was obtained for one- and two-component signals and the effects of diotic interference were assessed with two-, three-, and six-component signals. For all signals, the dichotic components contained interaural differences in time (whole-waveform delay), level, or both across a range of values. Listeners with higher audiometric thresholds performed more poorly in all conditions, despite the use of equivalent sensation levels. Diotic interferers reduced performance, but there were no additional impacts of age or absolute threshold. Statistical analyses indicated that the impacts of age were primarily due to the co-occurring increases in pure-tone thresholds, suggesting that: (1) absolute threshold is a better predictor of binaural ability than age and (2) reduced sensitivity to binaural differences is not addressed by presenting stimuli at equivalent sensation levels. [Work supported by VA RRD CDA-2 C4963W.]

4pPP19. Age-related differences in the time it takes to form auditory images of broadband noises. Mengyuan Wang (Dept. of Psych., Human Commun. Lab., Univ. of Toronto at Mississauga, 3359 Mississauga Rd. North, Toronto, ON L5L 1C6, Canada, mengyuan.wang.pku@gmail.com), Bruce Schneider (Univ. of Toronto at Mississauga, Toronto, ON L5L 1C6, Canada), Yanhong Wu, Xihong Wu, and Liang Li (Peking Univ., Beijing, 100871, China)

When correlated noises are presented over earphones to the two ears, listeners typically form a fused compact image of the noise. However, when the noises presented to the two ears are independent, listeners tend to hear two noises: one on the left and the other on the right. In Experiment 1 we determined the minimum duration required to form auditory images by asking listeners to distinguish between a 1-s presentation of independent noises to the left and right ears, and another 1-s presentation in which the noise was correlated for x ms before switching to two independent noises. In Experiment 2, one of the noises was correlated throughout the 1-s presentation; the other started off uncorrelated before switching to correlated. Younger adults performed better than older adults in both experiments. However, the performance of younger adults was better in Experiment 2 than in Experiment 1, whereas the reverse was true for older adults. The implications of these results for age-related changes in auditory scene analysis will be discussed. [This work was supported by the National Natural Science Foundation of China, the Natural Sciences and Engineering Research Council of Canada, and the Canadian Institutes of Health Research.]

4pPP20. Temporal processing as a function of age: Interaural time difference discrimination. John H. Grose, Sara K. Mamo, Emily Buss, and Joseph W. Hall (Dept. OHNS, Univ. N. Carolina at Chapel Hill, 170 Manning Dr., Chapel Hill, NC 27599-7070, jhg@med.unc.edu)

Older listeners often exhibit auditory temporal processing deficits. This study tested for temporal deficits in presenescent hearing. Three age groups participated: younger (18–28 yr); middle-aged (40–55 yr); older (63–75 yr). All had normal lower frequency hearing (thresholds $\leq$ 20 dB HL, 250–2000 Hz). Exp. 1 measured ITD discrimination for tone frequencies from

250–1500 Hz (65 dB SPL). Stimuli were two 100-ms tone pulses (75-ms rise, 25-ms fall). The pulses were diotic in the standard stimulus; the signal had an ITD (random lead/lag) imposed on the trailing tone pulse. Older listeners generally had higher ITD thresholds than Younger listeners. Middle-aged and Younger listeners performed similarly at low frequencies but the middle-aged tended to hit ceiling (z radians) at lower frequencies. Exp. 2 measured discrimination of So and Sτ signals as a function of frequency. Stimuli were tonal carriers modulated at 5 Hz (4 periods). Standard AM tones were diotic; signal tones alternated between So and Sz during successive modulator periods. Middle-aged listeners often performed slightly poorer than younger listeners, but better than older listeners. These results suggest that temporal processing deficits are evident in the pre-senescent auditory system. [Work supported by NIDCD 5-R01-DC01507.]

4pPP21. Quality judgments for music signals by normal-hearing and hearing-impaired listeners. Kathryn H. Arehart (Dept. Speech Lang. and Hearing Sci., Univ. of Colorado, Boulder, CO 80309), James M. Kates (GN ReSound and Univ. of Colorado, Boulder, CO 80309), and Melinda C. Anderson (Univ. of Colorado, Boulder, CO 80309)

Noise, distortion, nonlinear signal-processing algorithms, and linear filtering can all affect the sound quality of a hearing aid or other audio device. Most hearing-aid research concentrates on speech, but music reproduction can also be an important factor in user satisfaction. In this presentation, quality judgments are made for several different music signals by normal-hearing and hearing-impaired listeners. The music signals include orchestral classical music, jazz instrumental, and vocal. The signal processing uses a simulated hearing aid. The noise and nonlinear signal degradations include additive noise, multitalker babble, peak-clipping distortion, quantization noise, multichannel compression, spectral subtraction. Linear filtering conditions include bandwidth limitation, spectral resonance peaks, and spectral tilt. Conditions combining noise and nonlinear processing with linear filtering are also included. Subject ratings of the degraded music will be presented, along with comparisons of the ratings of music with that of speech for the same set of processing conditions. In addition, the subject ratings will be compared predictions using quality metrics based on auditory models.

4pPP22. A speech quality metric based on a cochlear model. James M. Kates (GN ReSound and Univ. of Colorado, Boulder, CO) and Kathryn H. Arehart (Univ. of Colorado, Boulder, CO)

Noise, distortion, nonlinear signal-processing algorithms, and linear filtering can all affect the sound quality of a hearing aid. The sound quality, in turn, can have a strong impact on the success of the device. The general approach in designing a quality metric is to compare the degraded signal with a clean (unprocessed) reference signal; the comparison generally involves comparing one or more features extracted from the signals. In this presentation, features are extracted from a computationally efficient model of the auditory periphery. The model includes the middle ear, an auditory filter bank, dynamic-range compression, two-tone suppression, and loudness scaling. The first step is a metric for noise, distortion, and nonlinear signal processing. The noise and nonlinear metric focuses on differences in the short-time signal behavior. The second step is a metric that focuses on the changes in the long-term average spectrum caused by linear filtering. The third and final step is to merge the noise and nonlinear metric with the linear filtering metric to produce a composite sound quality metric that can be applied to an arbitrary signal-processing system. The metrics give correlation coefficients better than 0.94 in comparison with quality ratings made by normal-hearing and hearing-impaired listeners.

4pPP23. Preference and performance for hearing-aid gain-compression prescriptions, based on a preliminary model of impaired sensitivity to tone intensity. Julius L. Goldstein (Hearing Emulations LLC, at Ariel Premium, 8825 Page Ave., St. Louis, MO 63114, julius@hearem.com), Michael Valente (Washington Univ., St. Louis, MO 63110), Metin Oz, and Peter Gilchrist (formerly at Hearing Emulations LLC)

A cochlearlike preliminary model was synthesized by joining a model of loss of cochlear "tip" gain, representing mild hearing loss, with a model for greater loss that translates the residual range for tone intensities into nonlinear "tail" expansion. Normal tip gain was modeled as 40 dB with square-root compression. Tail expansion was modeled with upper bounds for residual hearing at normal LDL of 100 dB HL, or at levels increasing by half decibel per decibel hearing loss above 40 dB; the latter provides less compression. A digital hearing aid comprising six octave-bandwidth (0.25–8 kHz) linear amplifiers with AGC was simulated to test alternative prescriptions with 32 experienced users of hearing aids. Preference for prescriptions with high or low compression and 4 or 8 kHz bandwidth was tested with clean speech repeated at input levels of 40, 60, and 80 dB SPL. The alternative best perceived as "mild," "comfortable," and "loud" was chosen; if uncertain, "either" was chosen. Performance was measured with intelligibility scores for low-probability speech sentences in noise (SPIN) at 8 dB signal to noise ratio and 70 dB SPL input. Independent of preference, high-compression prescriptions provided inferior performance, resulting from excessive increase in compression with hearing loss. [Work supported by NIDCD.]

4pPP24. A cochlearlike model of equal loudness levels for impaired and normal hearing: A new basis for hearing-aid gain-compression prescriptions. Julius L Goldstein (Hearing Emulations LLC, at Ariel Premium, 8825 Page Ave., St. Louis, MO 63114, julius@hearem.com)

Mean psychophysical data on most comfortable and uncomfortable tone levels as a function of hearing loss [Pascoe, 13th Danavox Symp. 153–183 (1988); D. M. Schwartz *et al.*, The Hearing J. **41**, 13–17 (1988)] is interpolated for other loudness levels by representing impaired auditory response as a modification of the idealized normal cochlear mechanical response to tones: linear at low and high sound levels joined by power-law compression of the low level "tip" gain. Denoting the low- and high-level nonlinear transitions as stimulus thresholds for compression and decompression, hearing loss elevates the compression threshold without changing the decompression threshold, until the compressive region is eliminated. Further hearing loss converts the compression threshold into a threshold for power-law "tail" expansion. Mean psychophysical data, represented by multiple regression lines, are fit exactly by an analytical solution for the normal model parameters. Estimates for thresholds of approximately 15 and 95 phons and tip gains of 40–52 dB are physiologically plausible for the normal cochlea. Normal equal loudness levels updated by Suzuki and Takeshima [Y. Suzuki and H. Takeshima, J. Acoust. Soc. Am. **116**, 918–933 (2004)] allow reliable extension of the model to low frequencies with estimates of auditory high-pass linear acoustic processing and of normal low-frequency tip gains. [Work supported by NIDCD.]

4pPP25. A weighting-function-based approach to subjectively modify the frequency response of a hearing aid. Andrew T. Sabin, Nicole Marrone, and Sumitrajit Dhar (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60201, a-sabin@northwestern.edu)

While approaches that modify the frequency response of a hearing aid based on a listener's subjective preference have demonstrated some success, they also have several limitations. Namely, these approaches (1) consider a small set of potential frequency-gain curves (FGCs) and (2) they converge on the final FGC, which makes the outcome dependent upon the starting point. Here we present a new approach that addresses these problems by (1) considering a much larger initial set of FGCs, and (2) using a method that is not convergence based. First, the listener rates the clarity of short samples of speech filtered by a set of maximally different probe FGCs. A weighting function is then computed, where the weight given to each ~1/3 octave frequency band is proportional to the normalized slope of the regression line between the listener's rating and the within-band gain of the probe FGC. Next, the listener rates the clarity of speech samples filtered by the weighting function, which is multiplied by one of several, randomly ordered, scaling coefficients. The final FGC is the scaling coefficient that receives the maximum rating, multiplied by the weighting function, added to a hearing-loss-specific correction factor. Results will be compared to current fitting strategies.

4pPP26. Effects of aging and irrelevant-dimension fluctuation on frequency doscrimination: Interference at late information-processing stages. Praveen Jajoria (Medical Sch., Univ. Texas Medical Branch, Galveston, TX 77555–0570) and Blas Espinoza-Varas (Dept. Commun. Sci. & Disord., Univ. Oklahoma Health Sci. Ctr., 1200 N. Stonewall, Oklahoma City, OK 73104)

Deficits in the ability to extricate relevant from irrelevant information at the perceptual level (i.e., selective attention) could help diagnose age-related cognitive decline at an early stage. Much experimental evidence supports this generalization, but there is debate as to where and how the irrelevant information interferes with the relevant one. To address this question, effects of irrelevant duration or level differences on pure-tone frequency discrimination thresholds (FDTs) were studied in healthy, normal-hearing, young and older adults. Target tones were presented either in isolation or followed by a distracter tone after a 90- or 350-ms silent interval; both tones were 1000 Hz, 80 ms, and 70 dB SPL. Irrelevant differences occurred simultaneously with relevant frequency differences or sequentially, in the distracter. With both presentation formats and silent intervals, FDT elevations (i.e., interference) were much larger in older than in young adults if the irrelevant differences were in level, but if they were in duration, FDT elevations were small in both groups. Since it occurs both simultaneously and retroactively (i.e., sequential), and depends on the similarity of relevant and irrelevant dimensions, the interference seems to take place at relatively late rather than early information-processing stages.

4pPP27. Dual-task costs in speech recognition by older and younger listeners. Karen S. Helfer, Jamie Chevalier, and Richard L. Freyman (Dept. of Commun. Disord., Univ. of Massachusetts, 358 N. Pleasant St., Amherst, MA 01003)

This study examined performance of younger and older adults when they carried out two auditory tasks simultaneously versus when performing each task in isolation. One task was to repeat a target sentence spoken by a male talker, presented simultaneously with two masking sentences spoken by two other male talkers. The second task was to judge whether neither, one, or both voices in the two-talker masker complex was presented time-reversed. Data were collected in both spatially-separated and colocated conditions. Results showed substantial individual variability within both subject groups as well as a strong effect of trial type order. Performance by most listeners declined to a greater extent for the natural/time-reversed task than for the speech recognition task during divided attention, even though subjects were instructed to give the two tasks equal weight. Importantly, requiring listeners to divide attention decreased or eliminated the spatial separation advantage for listening to speech in a competing speech situation. The poster will discuss differences noted between the two groups in overall performance as well as in the costs of dividing attention. [Work supported by NIH DC01625.]

4pPP28. Perceptually important acoustic features of environmental sounds. Laurie M. Heller, Benjamin Skerritt, and Emily Ammerman (Dept. of Cognit. and Linguistic Sci., Brown Univ., Box 1978, Providence, RI 02912, Laurie_Heller@brown.edu)

Listeners were asked to indicate the sources of dozens of sound events presented over headphones. The sounds were environmental sounds made by solid objects, liquids, and/or air undergoing a variety of actions taken from the website www.auditorylab.org. Each listener gave a series of responses about possible sources for each sound so that a similarity matrix could be constructed. The results of data reduction point to a manageable set of perceptual classes of sound events that can be described with a small number of distinctive spectro-temporal acoustic features. These features will be related to the attributes of the events that generated each class of sounds. Results for normal-hearing listeners will be analyzed in detail and preliminary results from hearing-impaired listeners will be previewed. [Work supported by NSF 0446955 and RI STAC.]

4pPP29. The incongruency advantage in elderly versus young normal-hearing listeners. Brian Gygi (Speech and Hearing Res., Veterans Affairs Northern California Health Care System, Martinez, CA) and Valeriy Shafiro (Rush Univ. Medical Ctr., Chicago, IL)

Previous research [Gygi & Shafiro (2007); Leech *et al.* (2007)] reported that environmental sounds heard in contextually incongruous naturalistic auditory scenes are better identified than sounds contextually congruous with the scene (e.g., rooster crowing in a farm ambience versus rooster in an office ambience). This incongruency advantage (IA) averages about four-to-five percentage points in p(c) and has been observed in both well-trained and naïve young normal hearing listeners [Gygi and Shafiro (2008)] and in children [Leech *et al.* (2007)]. One aspect of the IA is that it appears to be level-dependent, in that the effect is not present for lower sound/scene ratios (So/Sc). For experienced young listeners, the IA appears at a So/Sc of −15 dB. However, for naive young listeners, it is not manifest until −7.5 dB. A group of elderly normal-hearing listeners were tested and exhibited an IA only at So/Sc of −9 dB and above. The results show that the incongruency advantage are complex effects resulting from a combination of peripheral and central factors. [Research supported by the Merit Review Training Grant from the Dept. of Veterans Affairs Research Service, VA File # 06-12-00446, the National Institute for Aging, and the Rush Univ. Medical Center.]

THURSDAY AFTERNOON, 21 MAY 2009 FORUM SUITE, 1:25 TO 2:20 P.M.

Session 4pSAa

Structural Acoustics and Vibration: Vibro-Acoustic Diagnosis and Prognosis of Complex Structures II

Wen Li, Chair
Dept. of Mechanical Engineering, Wayne State Univ., Detroit, MI 48202

Invited Paper

1:25

4pSAa1. Synchronization of mechanical phase oscillators. David Mertens, Richard Weaver (Dept of Phys., Univ. of Illinois, 1110 W. Green St., Urbana, IL), and John Koliniski (Div. of Eng. and Appl. Sci., Harvard Univ., Cambridge, MA)

The Kuramoto model of a large number of weakly interacting phase oscillators exhibits a phase transition to synchronization. The literature has seen analytic theory and numerical simulations for both the Kuramoto model and sundry generalizations. Most real-world systems (e.g., flashing fireflies) imperfectly match the model, and their synchronization behaviors can therefore be taken to be in merely qualitative support of the theory. Here a mechanical system with well-understood microphysics is presented. A number N (of order 50) of eccentrically weighted DC motors (cell phone vibrators) is mounted on a plate. Each motor radiates into the plate and responds to its motion. If the plate dynamics is dominated by a single normal vibration mode the system is well described by finite-N Kuramoto

equations. Transitions to synchronization are observed in accord with theoretical expectations based on the quality factor of the plate, the number of motors, the ratio of the motor mass to the plate mass, and the natural speed of the motors. Calculations and results on a rich generalized-Kuramoto model are also presented, in which the plate dynamics entails many normal modes of vibration, and for which the governing equations resemble those of a laser.

Contributed Papers

1:50

4pSAa2. Analysis and prediction of underwater sound radiation from a cylindrical pile driven by an impact hammer. Mardi C. Hastings (Appl. Res. Lab., Penn State Univ., P. O. Box 30, State College, PA 16804, mch26 @psu.edu)

A vibroacoustic model for sound radiation from a submerged pile was developed to predict sound pressure levels generated during offshore construction activities. It was implemented in MATLAB and verified with measured data reported in the literature. The modal response of a cylindrical pile to hammer impact is first determined. Contributions from all dominant modes are then summed at various ranges to produce a two-dimensional spatial mapping of sound pressure level. The analysis was used to investigate the effects of pile material, size, and geometry on radiated sound pressure levels. Results indicate that smaller diameter piles of the same material do not necessarily produce lower sound pressure levels, but rather shift the peak sound pressure levels to higher frequencies. Because these basic parameters affect bandwidth as well as sound levels, there is no simple way to classify source level and frequency. Peak sound pressure levels generated underwater by driving steel piles with impact hammers into the sea bottom can easily exceed 200 dB re 1 μPa at a range of 10 m, so accurate prediction of the sound field is needed to assess potential risks to the environment.

2:05

4pSAa3. On vibroacoustic modal analysis of arbitrarily shaped vibrating structures. Huancai Lu (SeeSound 360, 3649 Glenwood Ave., Windsor, ON N9E 2Y6, Canada, huancai.lu@gmail.com) and Sean Wu (Wayne State Univ., Detroit, MI 48202)

Vibroacoustic modal analysis based on Helmholtz least squares method (HELS) is presented in this paper to explore the inherent dynamic characteristics and correlation of structural vibration and sound radiation of arbitrarily shaped vibrating structures. A series of mutually orthogonal vibroacoustic components is established by decomposing the normal velocity and normal acoustic intensity, which are reconstructed on source surface using HELS, in an orthogonal space through singular value decomposition (SVD). By further analyzing the vibration efficiency and radiation efficiency, the contribution of each of individual modal components to the resultant structural vibration and sound radiation can then be quantified. Therefore one is allowed to identify the vibroacoustic modal components that are most responsible for resultant structural vibration and sound radiation. The test object is a clamped and baffled thin rectangular steel plate, which was excited by a shaker and tested in a semianechoic chamber. The reconstructed normal surface velocities of the plate are validated against the results scanned by laser equipment. The vibration efficiency and radiation efficiency of the vibroacoustic modal components are examined, and the correlation between structural vibration and sound radiation of components at several frequencies is analyzed.

THURSDAY AFTERNOON, 21 MAY 2009 FORUM SUITE, 2:25 TO 3:55 P.M.

Session 4pSAb

Structural Acoustics and Vibration and Engineering Acoustics: Concepts of New Vibration Sensors

Daniel W. Warren, Chair
Knowles Electronics Inc., 1151 Maplewood Dr., Itasca, IL 60143

Invited Papers

2:25

4pSAb1. Role of slowness mapping in determining the directions of acoustic and seismic signals. Qamar A. Shams, George R. Weistroffer, John W. Stoughton (NASA Langley Res. Ctr., M.S. 238, Hampton, VA 23681), and Allan J. Zuckerwar (Analytical Services Mater., Hampton, VA 23666)

Slowness mapping is a method to estimate the angle of arrival of plane waves propagating across a sensor array. A review of time-delay estimation and its application to slowness vector estimation, the forward model, the inverse model, azimuth estimation, and elevation estimation will be presented. A method for performance grading with "out-of-bounds" conditions is described, and in the special case of subsurface acoustic sensors, a method for discriminating against seismic signals. The method has been applied to locate the direction of Space Shuttle and other rocket launches, infrasonic emissions from clear air turbulence, and incidental sources found in the environment.

2:50

4pSAb2. Experimental validation of alternate integral-formulation method for predicting acoustic radiation based on particle velocity measurements. Zhi Ni and Sean Wu (Wayne State Univ., 5050 Anthony Wayne Dr. Detroit, MI 48202)

This paper presents experimental validation of an alternate integral-formulation method (AIM) for predicting acoustic radiation from an arbitrary structure based on the particle velocities measured by a laser Doppler anemometer (LDA) and double hotwire sensor over a hypothetical surface enclosing a target source. Both the normal and tangential components of the particle velocity on this hypothetical

surface are measured and taken as input data to AIM codes to predict acoustic pressures in the exterior region. The results thus obtained are compared with those measured by microphones at the same locations. Measurement limitations are discussed and an error analysis of LDA measurement is presented.

3:15

4pSAb3. Visualization of vibrations measured with a multi-channel optical vibration sensor. Jun Hasegawa and Kenji Kobayashi (Dept. of Electron. and Comp. Syst., Takushoku Univ., 815-1, Tatemachi, Hachioji-shi, Tokyo 193-0985, Japan, jhase@es.takushoku-u.ac.jp)

Displacement-type sensor units made of optical fibers were developed to realize multi-points measurement of vibrations with high spatial resolution. Each sensor unit has the displacement resolution of 10 nm within the dynamic range of more than 90 dB, the frequency bandwidth of up to 80 kHz. Up to 64 sensor units can be arrayed as a multi-channel sensor head, with the minimum gap between sensor units of 4 mm. It means that the spatial resolution for the multi-points measurement is 4 mm. The calibrating method with the measuring object was developed to realize accurate measurements. During the calibration phase, the object is kept stand-still, and only the sensor head moves by the linear actuator. Thus, the calibration data can be obtained just before the measurement. Several arrangements of the sensor system is available depending on the object to be measured, such as line vibration, surface vibration, and the 3D movement of small object. The developed system has been used for the measurement of the actuators for vibratory microinjection and the measurement of artificial heart valves. Those results showed the advantage of the system. [This study has been supported by JSPS under the Grants-in-Aid for Scientific Research, Nos. 14550427 and 16560369.]

Contributed Paper

3:40

4pSAb4. Acoustic sensor structural configuration study. Bill B.B. Zhang and W. Steve Shepard, Jr. (Dept. of Mech. Eng., The Univ. of Alabama, Box 870276, Tuscaloosa, AL 35487, sshepard@eng.ua.edu)

This work examines the configuration requirements for a sensor being developed to provide information about the location of an acoustic source. The continuous sensor's vibrational response, caused by a traveling acoustic wave, is used in an inverse method to reconstruct the forces in that wave. A finite-length beam supported on an elastic foundation is the candidate structure. The impact of changing various structural parameters on the beam's vibration response is examined analytically using finite element methods. A goal in configuring the beam sensor is to obtain a response with enough dynamic information to accommodate a successful force reconstruction. Some of the configuration parameters of interest include the beam material, geometry, thickness, and the elastic foundation properties. The excitations used in the study consist of transient waves with differing numbers of sinusoidal half-cycles. Because background noise has the potential to impact the force reconstruction, various amounts of white noise are added to the simulated response values prior to performing the traveling wave reconstruction. Tikhonov regularization and the L-curve method are used. Results obtained from the simulations show that this sensor model is effective in identifying moving wave loads. [Work supported by NSF Sensor Innovation and Systems.]

THURSDAY AFTERNOON, 21 MAY 2009 GRAND BALLROOM I, 1:00 TO 4:00 P.M.

Session 4pSC

Speech Communication: Second Language and Cross-Linguistic Speech (Poster Session)

Megha Sundara, Chair
Dept. of Linguisitcs, Univ. of California at Los Angeles, Los Angeles, CA 90095-1543

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 even-numbered papers will be at their posters from 2:30 p.m. to 4:00 p.m.

4pSC1. An acoustical comparison of English tense and lax vowels. Byunggon Yang (Dept. of English Education., Pusan Natl. Univ., 30 Changjundong Keumjunggu, Pusan 609-735, South Korea, bgyang@pusan.ac.kr)

Several studies on the pronunciation of English vowels point out that Korean learners have difficulty producing English tense and lax vowel pairs. The acoustic comparisons of those studies are mostly based on the formant measurement at one time point of a given vowel section. However, the English lax vowels usually show dynamic changes across their syllable peaks and subjects' English levels account for various conflicting results. The purposes of this paper are to compare the temporal duration and dynamic formant tracks of English tense and lax vowel pairs produced by five Korean and five American males. Results showed that both the Korean and American males produced the vowels with comparable durations. The duration of the front tense-lax vowel pair was longer than that of the back vowel pair. From the formant track comparisons, the American males produced the tense and lax pairs much more distinctly than the Korean males. The results suggest that the Korean males should pay attention to the F1 and F2 movements, i.e., the jaw and tongue movements, in order to match those of the American males. Further studies are recommended on the auditorily acceptable ranges of F2 variation for the lax vowels.

4pSC2. The production of new versus similar vowels by Korean-English bilingual children. Soyoung Lee (P.O. Box 413, Enderis Hall 855, Univ. of Wisconsin-Milwaukee, 2400 E Hartford Ave., Milwaukee, WI 53201) and Gregory Iverson (Univ. of Maryland, College Park, MD 20742-5121)

The purpose of this study was to compare the vowel productions of Korean-English bilingual children with those of monolingual English and Korean-speaking children. Flege (1987) hypothesized that equivalent clas-

sification prevents adult L2 learners from establishing separate phonetic categories for similar phones (which occur in both languages) but not for new L2 phones. This hypothesis, however, has not been fully tested in bilingual children. The present study examined twenty 5-year-old Korean-English bilingual children who had been exposed to both languages for at least 2 years along with their age-equivalent monolingual counterparts, for a total of 60 children. Nine English vowels and six Korean vowels were elicited using picture cards. First and second formant frequencies for similar (/i, ɛ, u, o, a/) and new (/ɪ, ɔ, ɨ, æ, i/) vowels were measured. Vowels which were categorized as new were not appreciably different from those of monolinguals, suggesting that bilingual children produce new vowels authentically, consistent with the Flege hypothesis. However, the findings were mixed with respect to similar vowels. Bilingual children produced English vowels similarly to monolingual English-speakers, but their Korean vowels differed from those of monolingual Korean children, implying influence of the dominant (English) over the less-dominant language (Korean).

4pSC3. Acoustic features and inteilligibility of American-English vowels for English, Chinese, and Korean talkers. Su-Hyun Jin, Chang Liu, and Sangeeta Kamdar (Dept. of CSD, Univ. of Texas Austin, 1 University Station, Austin, TX 78712)

Sixteen American-English vowels including 12 monothongs and 4 diphthongs were recorded in a phonetic context of /hVd/ from young English (E), Chinese (C), and Korean (K) talkers. The Chinese and Korean talkers were bilingual and had stayed in United States up to 6 years. Two sets of experiments will be discussed: acoustic analysis and intelligibility of English vowels produced by the three groups of talkers. Results of acoustic analysis showed that there was no significant difference in F1×F2 vowel space among the three groups of talkers. In addition, the three groups of talkers showed great similarity in F2/F1 ratio across the 12 monothongs. Vowel durations had significantly greater variability across vowel categories for the Chinese and Korean talkers than for the English talkers, indicating that, besides producing spectral differences among vowels, Chinese and Korean talkers also attempted to generate durational difference among vowels to make each vowel distinguishable from others. More acoustic features such as spectral tilt and formant transition in the diphthongs and the effects of acoustic features on vowel perception by native English listeners will be discussed. Furthermore, the relationship between the vowel intelligibility and the second language experience of non-native talkers will be examined.

4pSC4. How regional dialect effects second language learning. Mu-Ling Teng (No. 1001, Univ. Rd., Hsinchu, Dept. of Foreign Lang. and Lits, Natl. Chiao Tung Univ., Taiwan 300, ROC), Yi-Chu Ke, Chu-ting Chen, Bo-hong Lu, Lai-Iok Ip, Ho-hsien Pan (Natl. Chiao Tung Univ., Taiwan), Shih-wen Chen (Natl. Tsing Hua Univ., Taiwan), and Hsiu-Min Yu (Chung Hua Univ., Taiwan)

This paper discussed dialect effects on L2 learning. Data from advanced English-learning students were separated into groups according to geographic regions: Northern, Middle, and Southern parts of Taiwan. Subjects were asked to read out loud eight English words, namely, "heed," "hid," "head," "had," "hod," "hawed," "hood," and "who'd." Acoustic information, such as F1 and F2, were measured; also, since Taiwanese English learners tend to use temporal cues to distinguish tense and lax vowels, duration of each word were also measured. Comparing the vowels read by the subjects with native speakers, we found that the Northern group had very similar patterns with native speakers in both spectral and temporal cues. As for the Middle group, the vowel shape was going upper right in the vowel chart and the duration pattern was quite distinctive from native speakers. Furthermore, all the subjects from the Southern part of Taiwan speak both Mandarin and Taiwan Min and their vowel spaces were the most deviated from native speakers, from which we assumed that PAM model (Best, 1993, 1995) could give a reasonable explanation: they have two vowel systems that could be referenced from, thus they might easily get confused. However, several problems still need further examination.

4pSC5. Production of prosodic cues by Beijing Mandarin speakers in second language (L2) English. Karen A. Barto-Sisamout (SLAT Office, P.O. Box 210014, Univ. of Arizona, Tucson, AZ 85721, kabarto@email.arizona.edu)

Does the prosody of speakers' first language (L1) influence their prosody in their second language (L2)? The current work investigates this topic for a tone language, Beijing Mandarin (BMan), as L1, and an intonation/stress language, English, as L2. English uses F0 contours as part of the intonation system, to signal syllabic prominence in a word, word prominence in a phrase, and the difference between questions and statements. In English there is a pitch peak delay, where the F0 peak occurs after the stressed syllable in two-syllable stress-initial words. Conversely, BMan uses F0 contours lexically, as a lexical tone language, and the F0 peak is on the stressed syllable. The production of these F0-related prosodic cues is investigated with BMan and Native English production of English narrow and broad focus statements and questions, with measurements of F0 contours on accented and unaccented words, pitch peak delay on accented words, and F0 levels in statements and questions. The goal is to learn if BMan speakers lexically specify tone of certain word types in English like in their L1, or realize F0 contours intonationally, like English speakers, or, finally, if they use an intermediate system different from L1 and L2.

4pSC6. Limited effects of early language learning on second language speech production. Tetsuo Harada (School of Education, Waseda Univ., 1-6-1 Nishi Waseda, Shinjuku, Tokyo 169-8050, Japan, tharada@waseda.jp)

Knightly *et al.* (2003) report long-term production benefits of childhood second language (L2) exposure in a naturalistic setting. However, it is unknown whether or not their finding will apply to classroom L2 learning. This study compares the production of voice onset time (VOT) in English and Japanese and closure duration of singletons and geminates in Japanese by English-speaking university students ($n=15$) who graduated from a Japanese immersion program (early learners) and adult learners of Japanese ($n=15$) with no exposure to Japanese in childhood (late learners). The informants were asked to repeat several target words including initial /p, t, k/ for VOT, and medial /p, t, k/ and /pp, tt, kk/ for singletons and geminates in a sentence frame. Both VOT of the initial stops and closure duration of the medial stops were measured. The results show that the production of VOT and closure duration by the early learners was not significantly different from that of the late learners ($p=0.71$ and $p=0.13$, respectively). The findings may hypothesize that there are only limited effects of classroom L2 exposure in childhood on L2 speech production in adulthood. [Work supported by Grant-in-Aid for Scientific Research (C) 20520527.]

4pSC7. The effect of oral proficiency on production of rhythm in spontaneous second language (L2) Japanese speech. Irina Shport (Dept. of Linguist., Univ. of Oregon, Eugene, OR 97403, ishport@uoregon.edu)

This study addresses the question of whether oral proficiency in Japanese second language (L2) speech has a unique correlation with acoustic characteristics of rhythm production that is independent from segments. Among four rhythm measures used (V%, VarcoV, VarcoC, VI-M), only two measures were different for the spontaneous L2 Japanese speech of beginning and intermediate learners. The interlanguage rhythm of less proficient speakers of Japanese was characterized by lower variability in duration of vocalic stretches (VarcoV) and higher variability in duration of consonantal stretches (VarcoC), $p < 0.05$. For both VarcoV and VarcoC values, the distribution of the individual speakers' rhythm scores was much tighter and on target for the intermediate students than for the beginning students. Furthermore, VarcoV values were significantly correlated with number of utterance-final vowels, and VarcoC values were correlated with number of obstruent clusters. In sum, the findings suggest that rhythmic differences in spontaneous L2 speech have an epiphenomenal nature stemming from the segmental structure of Japanese: the acquisition of mora-timed rhythm by learners of Japanese seems to be contingent upon the target-like production of segments which varies with proficiency level of learners.

4pSC8. Cross-linguistic differences in prosodic organization: Evidence from a repetition task. Emily Nava and Louis Goldstein (Dept. of Linguist., Univ. of Southern California, GFS 301, Los Angeles, CA 90089)

The prosody of languages such as English and Spanish has been characterized as exhibiting different rhythmic organizations. English has been hypothesized to organize syllables into feet, with one stressed syllable per foot. Spanish is among the languages whose prosody has been hypothesized to not include the foot, despite the existence of lexical stress in Spanish. We tested the hypothesis that these potential differences in organization would

lead to systematically different responses when speakers were asked to entrain a sequence of syllables with a metronome at an increasing rate, and that L1Spanish/L2English speakers would continue to exhibit the Spanish pattern in their English. Speakers of all three types were recorded producing a single repeated syllable, or a sequence of two alternating syllables, in time with a metronome, whose rate increased monotonically after a stabilization period. At slower speech rates, English speakers produced each word as a separate foot with a corresponding pitch accent, while at increased speech rates, two words were grouped together into a single iambic foot with one pitch accent. Spanish speakers and L1Spanish/L2English speakers are expected to shorten both vowels, employing a symmetric strategy (as opposed to the asymmetric strategy of English speakers) to keep pace with the metronome. [Work supported by NIH, NSF.]

4pSC9. Effect of word length on vowel production by Mandarin and American speakers: Comparison of [i] and [ɪ]. Chung-Lin Yang (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405, cy1@indiana.edu)

Yang (2008) found that Mandarin speakers' productions of [eɪ] and [ɛ] (as in *gate-get*) showed much less difference than Americans made. In addition, they made a smaller distinction in disyllables than in monosyllables. The current study compares the production of [i] and [ɪ] (*beat-bit*) in the same conditions. Target-vowels were embedded in Stop-V-Stop context in carrier sentences. Vowel durations were measured after stop release until final closure. Formant measurements were made at two time-points for each token. Preliminary analysis extends previous results of reduced contrast between tense and lax vowels by the Mandarin speakers to vowels [i]-[ɪ] and also shows greater reduction of the contrast in disyllabic words relative to monosyllables. The Mandarin speakers' formants for [i]-[ɪ] were almost identical in the disyllabic words, and their duration ratios were not significantly different, whereas American speakers' measurements were very different between two vowels and the same in disyllables as monosyllables. The current study provides further evidence for an effect of word length on English vowel contrasts by Mandarin speakers.

4pSC10. Short-term cross-linguistic interactions in bilinguals' vowel production. Matthew T. Carlson, Howard Nusbaum, and Shannon L. Heald (Dept. of Psychol., Univ. of Chicago, 5848 S. Univ. Ave., Chicago, IL 60637, carlsonmt@uchicago.edu)

Evidence suggests that the languages of a bilingual interact, and research has examined how first language (L1) phonetic categories influence second language (L2) acquisition. Although there may be substantial cross-linguistic interaction in the early stages of L2 learning, less is known about acoustic-phonetic variability when the bilingual's systems are more stable. This study examines L1 and L2 phonetic interaction, testing how the production of vowels in one language affects vowel production in the other. L1 and L2 vowel productions are compared for late English-Spanish bilinguals in an English-dominant environment. Five Spanish vowels and their English equivalents are examined. Subjects completed two experimental sessions, focusing respectively on L1 and L2, on separate days. Each session contained three blocks in which subjects produced vowels in isolation: two blocks in the target language separated by a block in the non-target language. Word cues were given to ensure the proper language for each block. Productions from first and third blocks of each session (i.e., in the same language) are compared to determine if intervening use of the other language affects subsequent vowel production. Implications are discussed for L1 and L2 phonetic representation, the status of a bilingual's languages, and second language acquisition.

4pSC11. Individual differences in the perception of vowels in a second language. A. Lengeris (Dept. of Speech Hearing and Phonetic Sci., Univ. College London, 2 Wakefield St., London WC1N 1PF, UK, a.leggeris@ucl.ac.uk)

Individuals may differ in their ability to learn the sounds of a second language (L2), but the origin of this variability remains uncertain. The present study examined whether individual differences in L2 vowel processing are related to individual differences in L1 vowel and/or nonspeech processing. Greek learners of English were given a large battery of perceptual tests examining their (1) identification of natural Greek vowels in noise, (2) identification and discrimination of synthetic Greek vowels in quiet, (3) identification of natural English vowels in quiet and in noise, (4) identification and discrimination of synthetic English vowels in quiet, and (5) discrimination of a nonspeech (F2 only) continuum in quiet. Preliminary results show a large degree of variability between individuals not only in L2 but also in L1 and nonspeech tasks. However, despite this variability, the participants were consistent in their performance across speech and nonspeech tasks, a finding that supports an auditory acuity rather than a speech-specific explanation for the individual differences in performance found in the data.

4pSC12. Perception of second-language stress and vowel quantity by English learners of Czech. Václav J. Podlipský (Dept. of English and American Studies, Palacký Univ., Krizkovskeho 10, 77180 Czech Rep., vaclavjonaspodlipsky@centrum.cz)

L2 acquisition may involve reattuning the perceptual system so that a cue becomes utilized for a new linguistic purpose. For instance, if vowel duration cues stress in L1 (like in English), whereas in L2 it marks phonological vowel quantity (like in Czech) it is of interest how the perception of stress and of vowel quantity will interact. Previous studies tested if stress affects vowel-quantity perception in such languages; this study explores conversely if vowel quantity affects stress perception in English learners of Czech. Since stress is word-initial in Czech, a word-boundary detection task was used, where stress either fell on a long vowel (stress and vowel length coincided) or on a short vowel, but there was a long vowel distracter in an adjacent syllable (stress and vowel length conflicted). Seventeen L1-English-L2-Czech and 69 Czech listeners were tested. Unexpectedly, natives scored slightly but significantly ($p<.05$) better when stress and vowel length conflicted than when they coincided. The opposite was true for the non-natives who were only at chance when stress and vowel length conflicted. No correlation between the non-native scores and various indices of L2 proficiency was found. It is concluded that learners transferred L1 perceptual strategies to the L2.

4pSC13. Identification and discrimination of tonal pitch in speech and nonspeech stimuli for Chinese- and English-native speakers. Chang Liu (Dept. of Commun. Sci. and Disord., The Univ. of Texas at Austin, 1 University Station A1100, Austin, TX 78712)

Given that Chinese is a tonal language while English is not, the present study is to examine how language background affects tonal identification and discrimination in speech and nonspeech signals for Mandarin-Chinese and American-English native speakers. The fundamental frequency (f0) contour was systematically manipulated from tone 2 (rising), to tone 1 (level), and to tone 4 (falling) for three types of signals: Chinese vowels /ɑ/, English vowels /æ/, and tonal sweeps. Both groups of listeners showed categorical perception in the identification of tones 1, 2, and 4, while Chinese-native speakers showed sharper boundaries across tonal categories than English-native speakers. Just-noticeable difference (JND) in tonal pitch was also measured for the standard tone 1, tone 2, and tone 4 of the three types of signals. JND was significantly smaller in speech stimuli for Chinese-native speakers than for English-native speakers, but quite comparable between the two groups of listeners for nonspeech stimuli. There results indicate that tonal pitch may be perceived specifically in speech stimuli, and for nonspeech stimuli, the capacity to process the tonal sweeps is similar across listeners with different language backgrounds.

4pSC14. Tonal confusion patterns in Mandarin by Cantonese listeners. Jung-yueh Tu (Dept. of Linguist., Simon Fraser Univ., 8888 Univ. Dr., V5A1S6 Canada, jta31@sfu.ca)

It is well-attested that linguistic experience affects the perception of non-native sounds. The vast majority of research on L2 perception has been carried out on segmental features from the perspective of phonetic similarity between the L1 and L2 sound systems. The general goal of this study is to expand our understanding of cross-linguistic comparison in non-native speech perception, focusing on the somewhat understudied area of perceptual assimilation at a suprasegmental level. More specifically, it deals with comparisons of prosodic systems in two tone languages, Mandarin and Cantonese. The current study examined perception of Mandarin tones by Cantonese listeners. In Experiment 1, Mandarin tones were presented and the Cantonese listeners were requested to identify which tone they heard. In Experiment 2, the Cantonese listeners were instructed to rate how similar each Mandarin tone was to a Cantonese tone. Preliminary results suggest that tonal confusion errors may result from not only the similar acoustic

properties of the tone pairs but also perceptual assimilation between L1 and L2 tonal contrasts. These findings are discussed in terms of the effects of L1 prosodic system on L2 perception and how perceptually assimilation patterns predict listeners' perception performance at the domain of lexical tones.

4pSC15. Acoustic-phonetic characteristics of naturally-elicited clear speech in British English. Rachel Baker and Valerie Hazan (Speech Hearing and Phonetic Sci., UCL, Chandler House, 2 Wakefield St, London WC1N 1PF, UK, rachel.baker@ucl.ac.uk)

Clear speaking styles are characterized by enhancements of specific acoustic-phonetic aspects of the speech signal. However, most studies of casual and clear speaking styles have been based on read speech recorded in laboratory conditions, both normally and when instructed to speak clearly. In this study, casual and clear speech produced by 40 talkers of southern British English was elicited in unscripted dialogues between two talkers via a series of "spot the difference" picture tasks, based on the "diapix" task developed by Bradlow and collaborators. In the "casual speech" condition, the two talkers conversed normally while doing the task. In the "clear speech" condition, to simulate communication between a normal-hearing talker and a talker with a cochlear implant, one of the talkers heard the other talker's speech passed through a three-channel vocoder. Additionally, talkers read sentences containing specific keywords that also occur in the diapix task, both normally and when instructed to speak clearly. Preliminary acoustic-phonetic analyses of the speech corpus will be presented. The clear speech elicited from the diapix and "read speech" conditions will be compared to ascertain whether clear read speech shows similar characteristics to clear speech elicited in more natural communicative situations. [Work supported by ESRC.]

4pSC16. Korean listeners' perception of consonant clusters in English. Gwanhi Yun (Dept. of English, Daegu Univ., 15 Naeri, Jinryang, Gyeongsan, Gyeongbuk, Korea 712-714, ghyun@daegu.ac.kr)

Perception experiments were conducted to see whether Korean listeners perceive the illusionary vowel in the legitimate sequence of English consonants, especially in onset positions. Korean listeners were presented three types of licit consonant clusters in English (e.g., sC in spee; Cy in kyoo; obstruent + approximant in plee) and had to choose between the original sequence and vowel-repaired sequence. First, Korean listeners did not perceive the illusionary inserted vowel between two consonants with all three types of onset clusters (sC, Cy, OA). Perception of the insertion of unmarked vowel was obtained only 20.1 %, while consonant clusters (disallowed in Korean) were not repaired in perception. Second, perceptual repair with illusive vowels was adopted the most in the sequence of obstruents and approximants (pree; 36 %), next of s+consonant (spee; 20%), and the least of Cy sequences (pyoo; 15%). Our result indicates that loanword phonology might be primarily computed on the basis of the phonological grammar of the borrowing language rather than perceptual assimilation based on phonetic-fine representation [Peperkamp *et al.* (2008)] . Further, it shows that perceptual repair by illusory vowels in the illicit onset clusters are gradual, not categorical unlike phonetic implementation of loanword phonology.

4pSC17. Improved learning of a non-native phonetic contrast by combining active task performance with passive stimulus exposures. Nicole Marrone and Beverly A. Wright (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208, n-marrone @northwestern.edu)

Adults can learn to hear the phonetic distinctions of another language with practice. It is commonly assumed that such improvements require active performance of the target task throughout training. However, we have seen that, for other perceptual tasks, similar performance gains can be achieved by replacing some active-performance trials with passive stimulus exposures. Here we examined the learning of a non-native phonetic contrast using several training protocols that varied in the proportion of active and passive trials during the practice period. Native, monolingual English speakers were trained to identify a Thai phonetic contrast along a voice-onset-time continuum. A group given half active training and half passive exposures showed as much improvement in the category boundary as a group given active training for the entire practice period. The active-passive group out-performed a group given half active training and half training on an unrelated task with no passive exposures. Finally, a group given only passive stimulus exposures showed the least improvement. Results suggest that learning a non-native phonetic contrast via active task performance can be reinforced by passive stimulus exposures. Therefore, active-passive training could be useful in real-world speech training applications. [Work supported by NIH/NIDCD.]

THURSDAY AFTERNOON, 21 MAY 2009 GALLERIA NORTH, 2:00 TO 5:00 P.M.

Session 4pSP

Signal Processing in Acoustics: Pattern Recognition in Acoustic Signal Processing II

Grace A. Clark, Chair
Electronics Engineering, Lawrence Livermore National Lab., Livermore, CA 94550

Chair's Introduction—2:00

Invited Papers

2:05

4pSP1. A tutorial on nonstationarity detection in acoustic signals: Parametric and nonparametric approaches. Patrick J. Wolfe (Statist. and Inf. Sci. Lab., Harvard Univ., Oxford St., Cambridge, MA 02138, patrick@seas.harvard.edu)

This tutorial provides an overview of nonstationarity detection in acoustic signals, focusing on model-based parametric approaches as well as more flexible nonparametric ones. The techniques discussed are presented in the context of speech and audio waveforms, with several real-world examples given, but also apply more broadly to any class of acoustic signals that exhibits locally stationary behavior. Many such waveforms, in particular information-carrying natural sound signals, exhibit a degree of controlled nonstationarity, and are often well modeled as slowly time-varying systems. The tutorial first describes the basic concepts of such systems and their analysis via local Fourier methods. Parametric approaches appropriate for speech are then introduced by way of time-varying linear predictive models, along with nonparametric approaches based on variation of time-localized estimates of the power spectral density of an observed random process. [Work supported in part by DARPA, NGA, and NSF.]

2:45

4pSP2. 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 B.C. V8W 3P6, Canada)

This paper considers matched-field tracking of a moving acoustic source in the ocean when acoustical properties of the environment (water column and seabed) are poorly known. The goal is not simply to estimate source locations, but to determine localization uncertainty distributions, thereby quantifying the information content of the tracking process, and to use this information to probabilistically predict future source locations. To this end, a Bayesian formulation is applied in which source and environmental parameters are considered random variables constrained by noisy acoustic data and by prior information on parameter values (e.g., physical limits for environmental properties) and on interparameter relationships (limits on source velocity). Source information is extracted from the posterior probability density (PPD) by integrating over unknown environmental parameters to obtain a time-ordered series of joint marginal probability surfaces over source range and depth. Given the strong nonlinearity of the localization problem, marginal PPDs are computed numerically using efficient Markov-chain Monte Carlo methods, including Metropolis-Hastings sampling over environmental parameters (rotated into principal components and applying linearized proposal distributions) and heat-bath Gibbs sampling over source locations. Several approaches to estimating optimal tracks, track uncertainties, and future source locations from the set of marginal-probability surfaces are considered.

3:25

4pSP3. Adaptation of tandem hidden Markov models for non-speech audio event detection. Mark Hasegawa-Johnson, Xiaodan Zhuang, Xi Zhou, Camille Goudeseune, and Thomas Huang (ECE Dept. and Beckman Inst., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801)

Non-speech audio event detection (AED) could be used for low-cost, spatially diffuse surveillance applications, e.g., monitoring of vehicle activity in a national park, or of footsteps in a hallway. Experiments have shown that non-speech AED benefits from the dynamic inference strategies such as the hidden Markov model (HMM), but that the acoustic features useful for non-speech events may not be the same as those useful for speech. One possible solution is a tandem HMM: an HMM whose observation vector is constructed from the output of an instantaneous discriminative classifier, e.g., a neural network. The use of tandem HMMs for non-speech AED is hindered, however, by the relatively small size of most non-speech-audio training corpora. This talk will demonstrate that tandem HMMs can be trained to detect non-speech audio events using a novel form of regularized training: Baum–Welch back-propagation (as proposed by Bengio *et al.*), using the conjugate-gradient adaptive form of the Baum–Welch auxiliary function (as proposed by Lee *et al.*, and as commonly used in maximum *a posteriori* HMM adaptation).

3:45

4pSP4. Statistical signal characterization in ocean acoustic tomography and geoacoustic inversions. Michael Taroudakis and Kostas Smaragdakis (Dept. of Mathematics, Univ. of Crete and FORTH/IACM, Knossou Ave, 71409 Heraklion, Greece, taroud@math.uoc.gr)

A performance study of a statistical characterization of an underwater acoustic signal in relation to geoacoustic inversion or tomography problems is presented. The method of characterization has been presented by Taroudakis *et al.* [JASA **119**, 1396–1405 (2006)] and is based on the use of an appropriate distribution law which describes the statistics of the sub-band coefficients of the signal wavelet transform. The method has been applied with synthetic data simulating tomographic experiments with low-frequency sources in shallow environments for the estimation of the water column and bottom properties. In this work the inversion procedure incorporating a genetic algorithm is applied in shallow water environments with simulated noisy data to assess the performance of the method and its limitations under realistic conditions.

4:05—4:15 Break

Contributed Papers

4:15

4pSP5. Wavelet-based neural networks applied to automatic detection of road surface conditions using tire noise from vehicles. Wuttiwat Kongrattanaprasert, Hideyuki Nomura, Tomoo Kamakura (Dept. of Elec. Eng., Univ. of Electro-Commun.s, 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan, wuttiwat@ew3.ee.uec.ac.jp), and Koji Ueda (Nagoya Electric Co., Ltd., Miwa-cho, Ama-gun, 490-1294, Japan)

The detection of road surface conditions is an important process in efficient road management. In particular, in snowy seasons, prior information about the road conditions such as an icy state, helps road users or automobile drivers to obviate serious traffic accidents. This paper proposes a novel approach for automatically detecting the states of the road surface from tire noises of vehicles. The method is based on a wavelet transform analysis, artificial neural networks, and the mathematical theory of evidence. The proposed method employs the wavelet transform using multiresolution signal decomposition techniques. The proposed classification is carried out in sets of multiple neural networks using learning vector quantization networks. The outcomes of the networks are then integrated using the voting decision-making scheme. It seems then feasible to detect passively and readily the states of the surface, i.e., as a rule of thumb, the dry, wet, snowy, and slushy state, automatically.

4:30

4pSP6. Discrimination of blasting sounds, music, and wind noise using a Gaussian mixture-model classifier. D. Keith Wilson (Cold Regions Res. and Eng. Lab., U.S. Army Engineer Res. and Development Ctr., 72 Lyme Rd., Hanover, NH 03755, D.Keith.Wilson@usace.army.mil) and Michael J. White (Construction Eng. Res. Lab., U.S. Army Engineer Res. and Development Ctr., Champaign, IL 61822)

Wind noise and acoustic signals were recorded outdoors with a horizontally oriented, square, 7×7 microphone array having an intersensor spacing of 0.914 m. Even with this very small spacing, the wind noise was found to have very little spatial correlation at frequencies above 15 Hz, thus indicating that the turbulent eddies responsible for the wind noise distort very rapidly. This property can be used to distinguish wind noise from acoustic signals possessing similar frequency content. To demonstrate this idea, wavelet processing was applied to recordings with predominantly wind

noise, propane cannon blasts, and unsteady but persistent sounds (music). A set of features involving ratios of the wavelet coefficients at different frequencies and between different microphones was then formulated. Three of the features relate to the time-domain shape of the signals, two to the spectral content, and three to the spatial correlation. A Gaussian-mixture-model classifier was trained with the feature statistics of the three signal types, and classifier predictions were then compared against an independent test data set. Results indicate a 96.6% correct classification rate of the signals. Signal shape features reliably distinguish the blasting and music, whereas the coherence features reliably distinguish the wind noise from acoustic signals.

4:45

4pSP7. Audio enhancement of biomechanical impact data. Joe Guarino, Wes Orme, and Wayne Fischer (Dept. of Mech. and Biomedical Eng., Boise State Univ., Boise, ID 83725-2075, jguarino@boisestate.edu)

Analysis and interpretation of impact data from a force transducer or accelerometer can be augmented and enhanced using audio playback. Trends and differences which may be difficult to identify using data imagery can be elucidated and reinforced by converting digital data from a force plate to an audio signal, which can then be played through a high-quality speaker system. The audio stream can be processed using standard acoustical methods such as tempo and pitch shifting, which can emphasize frequencies and tone bursts for improved signal characterization. We apply audio enhancement to data from two separate biomechanical studies: (1) a drop landing experiment for the investigation of gender differences in impact upon landing and (2) an experiment for the investigation of impact differences between cleated and noncleated shoes on artificial turf. The continuous wavelet transform (CWT) is used to process data from a force plate in both studies. Results are compared using the semblance analysis approach described by Cooper and Cowan ["Comparing time series using wavelet-based semblance analysis," Computers Geosciences 34, 95–102 (2008)]. Interpretation of results is enhanced using speaker-driven audio output synthesized from the CWT and semblance analysis.

THURSDAY AFTERNOON, 21 MAY 2009 GRAND BALLROOM I, 1:00 TO 4:00 P.M.

Session 4pSW

Speech Workshop

Note: This is the first poster session scheduled for the Cross-Language Speech Workshop. Please see page 2751 for abstracts of the papers to be presented in this session.

THURSDAY AFTERNOON, 21 MAY 2009 BROADWAY I/II, 2:15 TO 5:30 P.M.

Session 4pUW

Underwater Acoustics and Structural Acoustics and Vibration: Monostatic and Bistatic Detection of Elastic Objects Near Boundaries: Methodologies, and Tradeoffs II

Mario Zampolli, Cochair
NATO Undersea Research Ctr., 19138 La Spezia, Italy

Karim Sabra, Cochair
Dept. of Mechanical Engineering, Georgia Inst. of Technology, Atlanta, GA 30332-0405

Invited Papers

2:15

4pUW1. Exact and approximate techniques for scattering from targets embedded in a layered medium. Ahmad T. Abawi and Michael B. Porter (HLS Res., Inc., La Jolla, CA 92037)

To be able to accurately compute scattering from a target embedded in a layered medium (waveguide), the scattering and propagation problems must be solved as a single boundary value problem. This is accomplished by solving the wave equation in an environment that contains both the target and the waveguide and satisfying boundary conditions on the surface of the target and the boundaries of the waveguide. One way that this can be accomplished is by the use of the virtual source technique, which replaces the target with a collection of sources whose amplitudes are determined from the boundary conditions on the surface of the target. This method converts the problem of scattering from a target in a waveguide to a multisource propagation problem. In this paper, the virtual source technique is used to compute scattering from a target in a waveguide and various ways to speed up computation are examined. These include the use of various propagation models to propagate the field produced by the virtual sources to the receiver. Various solutions are compared and the advantages and disadvantages of each model are discussed.

2:35

4pUW2. Bistatic specular reflection by a proud cone: Experiments and interpretation. Jon La Follett and Philip L. Marston (Phys. and Astronomy Dept., Washington State Univ., Pullman, WA 99164-2814, lafollej@mail.wsu.edu)

For sufficiently small grazing angles of illumination, specular backscattering of high frequency sound by a proud cone having a vertical symmetry axis is not observed from geometric considerations. Observations from scaled tank experiments suggest, however, that specular scattering from a conical target in this configuration can be observed at shallow grazing angles bistatically. Bright features in the scattering corresponded to sound that had interacted with the flat boundary and the cone and to direct acoustic reflections from the cone. Feature locations were in agreement with a geometric analysis of bistatic reflection [P. Marston, J. Acoust. Soc. Am. **124**, 2584 (2008)] modified so as to allow for images caused by the flat surface. Data were obtained by suspending the apex of a small solid aluminum cone through the air-water interface of a tank. The cone was illuminated from below at grazing incidence and the free surface simulated the ocean floor. Bistatic measurements were obtained by scanning a hydrophone along a line parallel to the interface at varied depths. Data were also obtained from a truncated cone. [Research supported by ONR.]

2:55

4pUW3. Excitation of low frequency modes of solid cylinders by evanescent waves: Mode properties and coupling. Aubrey L. Espana, Philip L. Marston (Dept. of Phys. and Astron., Washington State Univ., Pullman, WA 99164-2814), and Kevin L. Williams (Univ. of Washington, Seattle, WA 98105)

When using sound to detect objects buried beneath the seafloor, situations arise in which the incident acoustic wave has a significant evanescent component. This situation has been simulated in tank experiments [Osterhoudt *et al.*, IEEE J. Ocean. Eng. (in press)] and the simulation was used to investigate scattering mechanisms. In prior work, the backscattering of evanescent and ordinary-propagating waves from small solid aluminum cylinders was studied [Espana and Marston, J. Acoust. Soc. Am. **124**, 2584 (2008)]. It was determined that a strong low frequency mode could be excited when the cylinder was highly tilted in a horizontal plane. With increasing simulated burial depth, the spatial decay rate of the backscattering was enhanced compared to the spatial decay rate of the evanescent soundfield. This resonance has since been driven by evanescent waves when the cylinder is highly tilted in a vertical plane. This alternate method of excitation also showed an enhanced spatial decay rate with increasing simulated burial depth. FEM simulation of the free-field response of the cylinder reveals that this mode causes a rocking motion of each end of the cylinder and it is plausible that evanescent waves could also excite this type of response. [Work supported by ONR.]

3:15—3:30 Break

3:30

4pUW4. Burial depth dependence of the bistatic scattering amplitude for cylinders illuminated by evanescent waves using two–dimensional finite elements. Nicholas R. Cerruti, David B. Thiessen, and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814, ncerruti@wsu.edu)

Prior research examined the dependence on simulated burial depth of the low-frequency scattering by small targets illuminated by evanescent waves [P. L. Marston, A. L. Espana, C. F. Osterhoudt, and D. B. Thiessen, J. Acoust. Soc. Am. **122**, 3034 (2007)]. The backscattering amplitude from targets with localized coupling displayed a spatial decay rate approximately twice that of the evanescent wave. An extended reciprocity relation was proposed which accounts for the more general case of a bistatic observation in the water column above the sea floor. In the bistatic case the spatial decay rate may differ from the case of backscattering. The present research concerns the testing of generalized reciprocity for small circular cylinders using two-dimensional finite elements. The calculated spatial decay rate for low-frequency bistatic scattering follows the generalized reciprocity condition when the predicted decay rate (for the specified observation scattering angle) exceeds the spatial decay rate of the incident evanescent wave. This computational result includes agreement with the double decay-rate case of backscattering. The calculations indicate that bistatic observation can significantly reduce the spatial decay rate of the signal dependence on burial depth. [Work supported by ONR.]

3:50

4pUW5. Bistatic and monostatic scattering from elastic structures using boundary element methods in free space and near plane boundaries. Ralf Burgschweiger, Martin Ochmann (Beuth Hochschule fuer Technik Berlin, Luxemburger Str. 10, D-13353 Berlin, Germany, burgi@tfh-berlin.de), and Ingo Schaefer (Forschungsanstalt der Bundeswehr fuer Wasserschall und Geophysik, D-24148 Kiel, Germany)

The bistatic and monostatic numerical calculation of the pressure scattered from structures composed of elastic materials and possibly filled with another material is one of the main purposes for the detection of underwater objects. For this reason, the sound pressure scattered from spherical objects placed in and filled with fluid will be calculated in the frequency domain. The results of an in-house developed BEM-package which supports single and multiple fluid-structure-interactions will be compared with analytical solutions based on spherical wave functions and with results of commercial BEM/FEM applications. Performance optimizations of the calculation process for the uncoupled rigid and coupled monostatic case as a result of using a parallelized matrix creation and solution with a specific variant of the direct solving process will be presented. We will also compare results for a cubic structure, placed in water above a finite plane boundary, with an equivalent half-space solution that incorporates a suitable half-space Green's function and could be used for fast approximations in the mid- and high-frequency range.

4:10

4pUW6. Synthetic aperture sonar simulations of cylindrical targets. Steven G. Kargl, Kevin L. Williams, Eric I. Thorsos (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40 St., Seattle, WA 98105, kargl@apl.washington.edu), and Joseph L. Lopes (Naval Surface Warfare Ctr., Panama City, FL 32407)

Cylindrical targets of finite length can be used as reference targets not only for calibrating an existing SAS system, but more importantly, for testing new classification and identification algorithms. With only a few well-characterized measurements available for proud and buried cylindrical targets, numerical simulations of the acoustic response of these targets offer the potential to realize an unlimited set of target orientations with respect to the source and receiver locations. This paper discusses recent progress with our acoustic scattering models and the generation of a set of pings suitable for SAS processing. SAS images generated from numerical simulations are compared to SAS images generated from data collected during the recent pond experiment 2009 (Pondex09) at NSWCPCD's facility 383. The target is a solid aluminum cylinder with a 0.3 m diam and length of 0.61 m. [Work sponsored by ONR and SERDP.]

Contributed Papers

4:30

4pUW7. Wavefront curvature and near-field corrections for scattering by targets buried under sea-floor ripple. Raymond Lim, Gary S. Sammelmann, and Joseph L. Lopes (Naval Surface Warfare Ctr. Panama City Division, Panama City, FL 32407)

Curvature of the incident field wavefront has been implicated in discrepancies between modeled and measured pressure fields scattered at shallow grazing angles from targets buried under ripple in NSWCPCD's test pond, where the range to the target is limited to about 10 m. Improvements on our attempts to correct for these and other near-field effects in existing incident-plane-wave-based models is described here. Explicit numerical integration of incident field expansion coefficients as well as the basis functions needed to formulate the transition matrix solution for the scattering process is used in order to create a reliable benchmark for comparisons with simpler corrections and measured data. It is found that some care must be exercised in performing the required two-dimensional wave vector integrals because the ripple wave vector can cause integration to be unstable in the perturbative terms accounting for ripple effects. Results are compared with recent modifications of existing sonar simulations and data collected in NSWCPCD's test pond. [Work supported by ONR and SERDP.]

4:45

4pUW8. Bistatic scattering from underwater unexploded ordinance and the impact of burial. Joseph A. Bucaro (SET, Inc., Naval Res. Lab., 4555 Overlook Ave., Code 7130, Washington, DC 20375-5320, joseph.bucaro.ctr@nrl.navy.mil), Brian H. Houston (Naval Res. Lab., Washington, DC 20375), Larry Kraus (Global Strategies Group (North America), Crofton, MD 21114), Harry J. Simpson, David C. Calvo, and Louis R. Dragonette (Naval Res. Lab., Washington, DC 20375-5320)

Interest in exploring various sonar approaches for detecting underwater unexploded ordnance (UXO) has been growing in large part because of the strong support of the Strategic Environmental Research and Development Program (SERDP). Among the many issues now being explored are the following two fundamental questions: what are the broadband acoustic scattering characteristics associated with typical submerged UXO, and how are these impacted by the bottom sediment? Recently, the authors reported laboratory grade, *monostatic* acoustic scattering measurements on UXO targets in the free field. In the present study, further echo measurements are obtained on UXO both to infer what merits may exist for exploiting *bistatic* echo responses as well as to address the effect of sediment on the target echo characteristics. The measurements, which were carried out in the NRL sediment pool laboratory and free-field facilities, include bistatic *free-field* scattering for three source directions viz. 0°, 45°, and 90°, and bistatic measurements in the sediment facility with the target *proud, half-buried*, and *fully buried*. The data suggest that access to bistatic echo information in operations aimed at detecting submerged UXO-like targets could provide an important capability, especially for buried targets. [Work supported by SERDP and ONR.]

5:00

4pUW9. The broadband in-water structural acoustics of unexploded ordinance: Tank comparisons with at-sea rail measurements. Harry J. Simpson, Brian H. Houston, Mike L. Saniga, Joe A. Bucaro (Naval Res. Lab., Code 7130, 4555 Overlook Ave., Washington, DC 20375-5320), Alain R. Berdoz, and Larry A. Kraus (Global Strategies Group (North America) Inc., Largo, MD 2077)

Free field, proud, and buried laboratory measurements of unexploded ordinance (UXO) are compared to at-sea rail-based measurements of the same UXO. The Laboratory for Structural Acoustics (LSA) at the Naval Research Laboratory consists of a 1 million gallon, deionized water, indoor cylindrical tank (17 m diam by 15 m deep) and an indoor rectangular tank (10 m by 8 m) laboratory, with a 3 m deep sand bottom and 4 m of water column. These pristine laboratory measurements are used to explore the physical acoustics of the sound-structure interactions that can be used to validate models and to understand the structural acoustic features that may be measured in littoral environments. These laboratory results are compared and contrasted with measurements made on the same UXO in St. Andrews Bay near Panama City, Florida using a rail-based synthetic aperture sonar (SAS). The structural acoustics responses of the UXOs in this shallow water, 8 m deep, sandy-mud bottom environment are compared with the pristine results from the LSA. [Work supported by SERDP and ONR.]

5:15

4pUW10. Long-range low frequency shallow water acoustic propagation and bottom penetration experiments for underwater unexploded ordinance. Harry J. Simpson, Mike L. Saniga, Brian H. Houston (Naval Res. Lab., Code 7130, 4555 Overlook Ave., Washington, DC 20375-5320), Alain R. Berdoz, Philip Frank, Steve W. Liskey, and Larry A. Kraus (Global Strategies Group (North America) Inc., Largo, MD 20774)

Among the many issues now being explored in the acoustic detection underwater unexploded ordinance (UXO) in shallow water is the impact of propagation and bottom interaction. Here, we report on long range (1 km) propagation measurements in a littoral environment at the sediment-water interface in the 1 kHz–12 kHz frequency band. The water channel studied was an 8 m deep arm of St. Andrews Bay, Panama City, Florida. The medium grain sandy bottom had a compressional waves peed of 1700 m/s, and the wavespeed in the water was 1539 m/s. A dense array of point pressure measurements (1 cm vertical sampling) were acquired for a synthetic vertical aperture starting from 2 m above and going through 2 m below the interface and at three ranges, 78, 485, and 990 m. The results show deep penetration into the sediment at long ranges for frequencies below 5 kHz. The data are examined using a two-fluid parabolic equation model, and the overall environmental acoustics are discussed including the nature of sediment penetration due to evanescent waves, multipath, and interface roughness.

Session 5aAA

Architectural Acoustics: Architectural Acoustics Potpourri

Russell N. Altermatt, Chair
Altermatt Associates Inc., 522 SW Fifth Ave., Suite 1200, Portland, OR 97204

Contributed Papers

8:00

5aAA1. An application of the finite difference time-domain method to the study of coupled volumes. Jonathan M. Botts and Ning Xiang (Program in Architectural Acoust., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, bottsj@rpi.edu)

A finite difference time domain (FDTD) scheme is applied to study the behavior of sound energy decay in coupled-volume systems. This particular method takes into account frequency dependent wall impedance as well as source characteristics. In time-domain method, impulse responses are direct results of numerical calculation and transient effects specific to coupled volumes can be visualized. In addition to examining traditional acoustic metrics, this model allows for the investigation of the sound energy flows in the system. The numerical calculations will also be compared with experimental scale model measurements. Though limited to low frequencies due to computational load, this method will demonstrate many advantages over geometric acoustic models and other wave based models for the study of coupled volumes.

8:15

5aAA2. One-dimensional transport equation models for sound energy propagation in long spaces. Yun Jing (Graduate Program in Architectural Acoust., Sch. of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180), Edward Larsen (Univ. of Michigan, Ann Arbor, MI 48109), and Ning Xiang (Rensselaer Polytech. Inst., Troy, NY 12180)

In this paper, a three-dimensional transport equation model is first presented to describe the sound energy propagation in a long space. Then the three-dimensional model is reduced to a one-dimensional model by approximating the solution using the method of weighted residuals. The one-dimensional transport equation model directly describes the sound energy propagation in the long dimension and deals with the sound energy across the cross-section by prescribed functions. The one-dimensional transport equation model consists of a coupled set of N transport equations. Only $N = 1$ and $N = 2$ are discussed in this paper. For larger N, although the accuracy could be improved, the calculation time is expected to be significantly longer, which diminishes the advantage of the one-dimensional model over the three-dimensional model in terms of the computational efficiency. Some simulation results are provided in comparison with the ray-tracing based approach, which show good agreement.

8:30

5aAA3. Reciprocal maximum-length and related sequences in the generation of natural, spatial sounding reverberation. Uday Trivedi, Eric Dieckman, and Ning Xiang (Graduate Prog. in Architectural Acoust., School of Architecture, Rensseleaer Polytechnic Inst., Troy, NY 12180, triveu@rpi.edu)

The development of artificial reverberation has made considerable progress in recent years. In this research, a unique approach utilizing reciprocal maximum-length sequences and related sequences is taken in generating a natural, spatial sounding reverberation. Additionally, an algorithm to shape and vary spaciousness within the context of an artificially generated reverberation is implemented. Results are verified through experimental measurements and subjective testing methods in a multichannel format. This work sheds light on alternative methods of generating artificial reverberation outside of commercially available systems. As a result, a direct application to room modeling techniques can be drawn with this work. Extensions of this research will comprise of investigations in objective and subjective behaviors of double-sloped energy decays in acoustically coupled spaces. From an educational perspective, outcomes of this research can serve as a foundation in ear-training for acoustics students seeking a broadened understanding of perceived reverberation in varying contexts.

8:45

5aAA4. Optimized overdetermination for interior Helmholtz problem coupled solvers. Alexander Bockman, Paul Calamia, and Ning Xiang (Dept. of Architectural Acoust., RPI, 110 8th St., Troy, NY 12180, bockma@rpi.edu)

The use of points of overdetermination, so-called CHIEF points, in numerical solutions to exterior Helmholtz problems for the elimination of spurious modes is well established. Further, the number and relative position of such points in two-dimensional radially symmetric geometries has been demonstrated theoretically. Since CHIEF points require lower computational overhead than collocation nodes, strategic use represents an opportunity for the improvement of predictions with maximal efficiency. Given the differing uniqueness properties of solutions to the interior and exterior problems, the former allows for development of optimization in the absence of competing priorities. Theory for two- and three-dimensional analytical cases is developed and subsequently extended to general geometries. Localization with respect to subdomains of spatial decomposition is investigated as well.

9:00

5aAA5. Evaluation of the newly developed equation for noise barrier within an enclosed space. Lau Siu-Kit (Archit. Eng. Prog., 203C Peter Kiewit Inst., Univ. of Nebraska-Lincoln, 1110 South 67th St., Omaha, NE 68182-0681, sklau@engineer.com and) and Kwok-Wai Lam (The Hong Kong Polytechnic Univ., Hung Hom, Kowloon, Hong Kong)

Indoor barrier can be employed to ensure the privacy of speech in open-plan offices and to shield the noise generated by the machines in plant rooms. However, existing publications on diffraction and noise reduction by barrier inside an enclosed space using the classical theory is limited by the assumption of diffuse field. Recently, a new formula for estimating the barrier insertion loss inside an enclosed space was proposed by Lau and Tang [J. Sound Vib. **20**, 50–57 (2009)]. It is simple and is sufficiently accurate in engineering practice. In the present study, this new formula is proposed to estimate the insertion loss of a vertical partition between two flat rooms. The performance of the new formula is compared with that of the classical diffuse-field theory. The results indicate that the new formula provides more realistic and practical predictions of the insertion loss of the partition between two flat rooms than existing approaches. [The work described in this paper was supported by a grant from the Hong Kong Polytechnic University (Project A/C code: G-U362).]

9:15

5aAA6. Design and testing of a variable acoustics production studio with reflection free zone control room. Jon Mooney, Seth Harrison (KJWW Eng. Consultants, 623 26th Ave., Rock Island, IL 61201, mooneyjw@kjww.com), and Jon Brennan (Curtis, Inc., Cincinnati, OH 45203)

The new home of Curtis, Inc. in Cincinnati, OH, includes a recording studio and control room designed and tested by KJWW Engineering Consultants. The recording studio features adjustable wall treatments, which may be opened in varying degrees to adjust absorption and diffusion within the room. The adjoining control room has been designed to position the recording engineer inside a reflection free zone (RFZ). This design directs speaker reflections from the front of the room away from the recording engineer, extending the initial time delay (ITD) without requiring a dead front end. The control room also features unique wall and ceiling treatments, which serve not only as variable binary diffusers and absorbers but also as storage cabinets. These design elements allow for broad adjustment of the control room acoustics. The configurable design of the recording studio and control room coupled with exceptional sound isolation enable this facility to create flexible recording and listening environments. This paper will present an overview of: 1. The recording studio and control room designs; 2. The changes required during construction and their effects; 3. Acoustical measurements, quantifying the variable response of the rooms; 4. Operation of the rooms and preferred settings.

9:30

5aAA7. The acoustic shell of the Plaza de La Fe in Managua, Nicaragua. Russ Altermatt (Altermatt Assoc., Inc., 522 SW 5th Ave., S-1200, Portland, OR 97204)

The Plaza de La Fe, the Plaza to the Faith, of Pope John Paul II, in Managua, Nicaragua, is the site of many celebrations for the people of Nicaragua. At one end of the plaza, a performance, acoustic shell was constructed as a focus for performances in the plaza. The acoustic shell has a very large capacity, for full orchestra and choir, at a minimum. It supports multiple band setups on the same stage depending on the celebration requirements. The shell is used for all types of performances and presentations, be they multiband, dance, classical music, or speech. In spite of significant concavity of the overall design, the shell elements were also molded to provide directional sound reflections both to the plaza and to the performers on the stage. This presentation will show the progress of the project from design, through construction. The acoustic shell was inaugurated with an attendance of over 100,000 in Jan. 2005.

9:45

5aAA8. Classroom sound systems: Not a Green substitute for proper classroom acoustics. Richard F. Riedel (Riedel Audio & Acoust., 443 Potter Blvd., Brightwaters, NY 11718, riedelaudio@optonline.net)

The author will discuss the negative environmental impact that individual classroom sound systems create when used as a substitute for good acoustical design. School systems are being targeted by marketers of classroom sound systems and being convinced that they are a cost effective alternative to the standards that are put forth by ANSI S12.60-2002. This paper presents estimates of the potential waste that is created by these systems through future disposal of outdated technology and batteries used in these devices. The presentation will put forth suggestions for the dissemination of this information by interested parties to schools in their immediate vicinity.

10:00

5aAA9. Did paleolithic cave artists choose resonant locations for paintings? David Lubman (DL Acoust., 14301 Middletown Ln., Westminster, CA 92683, dlubman@dlacoustics.com)

Paleolithic artists were surely impressed by the acoustical properties of their cave dwellings. Since many inhabited caves and grottoes were highly reverberant, it is compelling to speculate that cave acoustics importantly influenced ancient artists. One imaginative suggestion is that Paleolithic artists intentionally chose the most resonant locations for painting. This thesis was proposed by Iegor Reznikoff and Michel Dauvois in the Bulletin de Ia Societe Prehistonque Francaise **85**, 238–246 (1988) and was engagingly described by British archaeologist Chris Scarre ["Painting by resonance," Nature (London) **338**, 382 (1989)]. Scarre found this thesis consistent with "the likely importance of music and singing in the rituals of our early ancestors." The researchers found a greater density of paintings where sonic resonances were judged strongest. However, that finding may be spurious: the same result is predictable from painterly rather than sonic considerations. Painters would likely choose nonporous rather than porous rocks to conserve pigment. Since nonporous rocks are generally highly sound reflective, nonporous rock locations tend to be highly resonant. Since porous rocks tend to absorb sound, porous rock locations are less resonant. Thus the higher density of cave paintings found at resonant locations may not imply the artists' preference for resonance.

10:15

5aAA10. Tehran concert hall—a virtual specimen indicative the conversion processes from architectural design to acoustical remedies. H. Azad (Dept. of Architecture, Camp. of Fine Arts, Univ. of Tehran, Enghelab st. Tehran, Iran, hassan.azad@gmail.com)

Over the last year of the B.Sc. period it was investigated by the author to design a 1200-seat concert hall as the thesis project. Now the quoted project has been enriched by the acoustical aspects of the concert hall design regarding the acquired capabilities during the M.Sc. studies. It has become prevalent these days to peruse and analyze the acoustical parameters of enclosures through the different acoustical simulation software such as Odeon 9, CATT-Acoustics 8, EASE 4.1, and Ramsete 2.5, either before or after their construction as an objective assessment but interpreting this data is a real challenge. [H. Azad. "Ali Qapu: Persian Historical Music Room?". Auditorium Acoustics Conference (2008), Oslo.] The variable height stage canopies taking into account their shape and materials is investigated by calculations with ray-tracing computer model to provide a large part of the audience with early reflections and to create good ensemble conditions for the musicians on stage. The diffusers design as a geometrical shape with design factors of height, surface coverage and orientation in addition to their location will be studied computer aided to improve the acoustics in this virtual concert halls.

Session 5aAB

Animal Bioacoustics and Signal Processing in Acoustics: Signal Processing Techniques for Subtle or Complex Acoustic Features of Animal Calls

Ann E. Bowles, Chair
Hubbs Sea World Research Inst., 2595 Ingraham St, San Diego, CA 92109

Chair's Introduction—7:30

Invited Papers

7:35

5aAB1. Distinguishing FM-modulated bowhead whale calls from airgun and other biologic signals using image processing, feature extraction, and neural networks. Aaron Thode, Delphine Mathias (Marine Physical Lab., Scripps Inst. of Oceanogr., 9500 Gilman Dr., San Diego, CA 92109-0238), Susanna Blackwell, Katherine Kim, and Charles Greene, Jr. (Greeneridge Sci. Inc., Goleta, CA 93117)

The detection and identification of repetitive or stereotyped bioacoustic signals in raw acoustic data are typically accomplished via matched filtering or spectrogram correlation techniques. However, the detection of FM-modulated sweeps that have variable frequency-modulated contours or bandwidths is still an active challenge. The problem is made more complex by the possible presence of other frequency-modulated signals such as airguns or other biologics. To provide a specific example, in 2007 and 2008 directional autonomous recording packages (DASARs) were deployed in the Beaufort Sea to monitor the annual migration of the bowhead whale (*Balaena mysticetus*) population during seismic exploration activities. A portion of these data sets has been manually analyzed, providing an opportunity to characterize the miss and false alarm rates of various FM-modulated detection and classification methods. Here literature and software on bioacoustic FM-contour tracing and classification are reviewed, with a particular emphasis on the use of contour tracing, image segmentation, feature extraction, and neural networks for identifying species-distinctive sounds. A combination of these methods, when applied to the Arctic data set, has been found to reduce the false alarm rate by a factor of 5, while preserving a miss rate of less than 20%. [Work supported by Shell Oil Co.].

7:55

5aAB2. A graph search algorithm for delphinid whistle contours. Marie A. Roch, Bhavesh Patel (Dept. of Comp. Sci., San Diego State Univ., 5500 Campanile Dr., San Diego, CA 92182-7720), Melissa S. Soldevilla, and John A. Hildebrand (Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093-0205)

A graph search algorithm is presented to extract tonal contours from audio signals in a fully automated manner. The algorithm is capable of tracking multiple contours simultaneously. Search is done in the spectral domain, with contours being treated as paths through a phase-magnitude space. A graph search formalism is combined with dynamic programming techniques to track candidate paths that may overlap or cross. Disambiguation of path crossings is accomplished by examining a combination of slope and the derivative of phase. The algorithm is general in nature and applicable to tonal calls from many animals with little modification. We target this work specifically toward delphinid whistles and demonstrate its efficacy on calls from bottlenose (*Tursiops truncatus*), short- and long-beaked common (*Delphinus delphis* and *D. capensis*), and Risso's (*Grampus griseus*) dolphins recorded in the Southern California Bight. Characterizing the complete contour will permit the development of call comparison techniques that take into account finer grained detail than the typically used contour statistics (e.g., number of inflection points, minimum, and maximum frequencies). [This work is sponsored by the Office of Naval Research.]

8:15

5aAB3. Noise-resistant acoustic measurements implemented in user-friendly software. David K. Mellinger (Cooperative Inst. for Marine Resources Studies, Oregon State Univ. and NOAA Pacific Marine Environ. Lab., 2030 SE Marine Sci. Dr., Newport, OR 97365)

Searching for vocalizations in lengthy marine recordings can be significantly enhanced by annotations indicating times, frequencies, and other characteristics of calls. Defining optimal search features, however, is difficult. For instance, one may wish to find harbor seal "roar" vocalizations, which extend up to 6 kHz, last 3–10 s, and have a broadband, non-tonal sound quality. What features best characterize such sounds? Marine recordings often contain vocalizations recorded at low signal-to-noise ratios, and it is essential that call measurements be consistent across different background noise levels. For instance, bandwidth is traditionally measured by manually indicating lower and upper frequency bounds in a spectrogram, and then subtracting the two. However, for vocalizations that fade at higher frequencies, like harbor seal roars, such bandwidth measurements can vary by a factor of three from low-noise to high-noise environments. Here we describe measurements, based principally on Fristrup's "Acoustat" approach, that have consistent values at variable noise levels. For instance, bandwidth is measured as the standard deviation of frequency, weighted by the per-frequency-bin normalized intensity. A set of noise-robust measurements have been implemented, including measures for duration, amplitude and frequency modulation, and many others. Implementation of these measurements in easy-to-use Ishmael and MATLAB software is described.

8:35

5aAB4. Signal processing challenges in the analysis of stereotyped killer whale calls produced in social contexts. Ann E. Bowles (Hubbs-SeaWorld Res. Inst., 2595 Ingraham St., San Diego, CA 92109, abowles@hswri.com), Jessica L. Crance, and Juliette S. Nash (Univ. of San Diego, San Diego, CA 92108)

Research on call development and usage by individual killer whales has been limited by signal processing challenges, which include (1) more than one simultaneously operating source, (2) a pulse-burst production mechanism, (3) a complex repertoire (7–17 calls), (4) gradation with nonstereotyped call types, (5) presence of nonstereotyped features encoding state (e.g., arousal), and (6) vocal matching among callers. To date, the most successful signal processing efforts have targeted tonal, modulated "screams." However, analysis of other component types (bursts, resting call), parsing of calls into components, and reliable classification of calls continue to require the intervention of a human brain. Under controlled laboratory conditions, using data from seven whales at SeaWorld San Diego, these characteristics were analyzed by ear and using traditional statistical methods. The work showed that important insights could be gained by comparing these call characteristics among individuals and social contexts. However, estimates of the sample sizes needed to show significant differences among individuals suggest that automation will be essential. Use of the human ear as a signal processor could also introduce biases. Efficient automation would greatly improve the study of individual call usage. [Work supported by the Hubbs-SeaWorld Society and SeaWorld San Diego.]

8:55

5aAB5. Feature vector selection for identifying killer whale individuals. Nicole Nichols, Les Atlas (Dept. of Elect. Eng., Univ. of Washington, Box 352500, Seattle, WA 98195, nmn3@u.washington.edu), and Ann Bowles (Hubbs-Seaworld Res. Inst., San Diego, CA 92109)

One hindrance to studying whales is the simple act of identifying individuals. The source filter model of human speech is a potentially useful model for killer whale vocalizations. With this assumption, evaluation of vocal model based optimizations for human speaker recognition was performed with a whale dataset. (Twenty one calls of type 1a or 1b, from four captive Icelandic killer whales.) Feature vectors were extracted from sound samples, which reduce dimensionality while encapsulating the identifying characteristics. Recognition accuracy is highly dependent on choosing a feature vector, which can accurately divide the sample space. Cepstral coefficients are the standard used for human speaker recognition. Modulation envelopes are known to be important for perception and have been shown to improve speech recognition performance [Kingsbury *et al.*, Speech Commun. **25**, 117–132 (1998)]. Modulation spectrograms were examined and sampled across acoustic frequency at the fundamental modulation rate of the call to extract modulation feature vectors. Visual analysis of modulation spectrograms note a possible pattern unique to individual whales. Presented herein is analysis of these patterns based on linear discriminant functions and hidden Markov models. [Work supported by National Science Foundation Graduate Research Fellowship, University of Washington, Hubbs-Seaworld Research Institute.]

Contributed Papers

9:15

5aAB6. Rhythmic analysis of sperm whale broadband acoustic signals. Natalia Sidorovskaia, Philip Schexnayder (Dept. of Phys., Univ. of Louisiana at Lafayette, UL Box 44210, Lafayette, LA 70504-4210, nas@louisiana.edu), Alexander Ekimov, James Sabatier (Natl. Ctr. for Phys. Acoust., University, MS 38677), George E. Ioup, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

Rhythmic analysis of sperm whale clicks, based on the static Fourier transform, was previously reported in the literature as a tool for extracting the acoustic signature of individual sperm whales [Andre and Kamminga, J. Mar. Biol. Ass. U.K. (2000)]. A new approach, originated from the algorithm developed for human motion analysis [Sabatier and Ekimov, "A Review of Human Signatures in Urban Environments Using Seismic and Acoustic Methods," Proc. IEEE (2008)], is applied to study time-dependent periodicity in sperm whale signals recorded in the Gulf of Mexico. The new approach for rhythmic analysis of sperm whales clicks, which are used for navigation purposes, shows the presence of stable temporal periodicity that could possibly be attributed to the acoustic portrait of individuals. [Research supported in part by SPAWAR.]

9:30

5aAB7. Sei whale localization and vocalization frequency sweep rate estimation during the New Jersey Shallow Water 2006 experiment. Arthur E. Newhall, Ying-Tsong Lin, James F. Lynch, and Mark F. Baumgartner (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543, anewhall@whoi.edu)

The Shallow Water 2006 (SW06) experiment was conducted in the Mid-Atlantic continental shelf off the New Jersey coast. A fast-sampling, 48-channel hydrophone array system recorded a number of sei whale (*Balaenoptera borealis*) vocalizations during this time. This system had 16 hydrophones on the vertical line array (VLA) component covering the water column from 13-m depth to the bottom (80 m) and 32 hydrophones on the 480-m horizontal line array (HLA) component that lay on the bottom. The sei whale receptions on the arrays are low-frequency (less then 100 Hz) downsweep chirps and have the typical acoustic modal arrival pattern seen in shallow-water, low-frequency sound propagation. Due to acoustic modal dispersion, the frequency sweep durations observed from the receptions are found to be longer than the original signal. A horizontal beamforming technique using the HLA component is implemented to determine the bearings to whales. The whale location along the determined bearing and the frequency sweep rate of original whale sound are simultaneously estimated using the VLA component with a multiple-parameter optimization technique. This optimization technique is based on acoustic normal mode theory and is designed to compensate for the effect of acoustic modal dispersion.

9:45—9:55 Break

Chair's Introduction—9:55–10:00

Invited Papers

10:00

5aAB8. The challenge of nonlinear dynamics in understanding mammalian vocalization. Michael J. Owren (Dept. of Psych., Georgia State Univ., PO Box 5010, Atlanta, GA 30302-5010, owren@gsu.edu)

Recent advances in vocal-fold physiology have shown that nonlinear dynamics play a central, but largely unexplored role in mammalian vocalization. Understanding that role is crucial to coming to grips with vocal signaling as a form of communication, as the "nonlinear phenomena" associated with coupled-oscillator systems like the larynx strongly shape their acoustic output. Those acoustics in turn are necessarily central to understanding both signaler and perceiver roles in the communication process. Issues that are fundamentally affected by the nonlinear nature of mammalian vocalization include categorizing sounds in acoustics and perception, breaking a repertoire into constituent call-types, understanding the identity cuing potential of various sound-types, and the relative roles of cognition and emotion in vocal production and perception. However, progress is being impeded by a general lack of accessible, quantitatively based, and objective tools to help in detecting, identifying, and measuring nonlinear phenomena in vocalizations. So far, researchers trying to include such events in their analyses have been limited to visual inspection of spectrograms, subjective classification, and simple counting. A pressing need thus exists for development of accessible, yet necessarily rather sophisticated tools for use in identifying and measuring occurrences of nonlinear phenomena in mammalian vocal acoustics.

10:20

5aAB9. Quantifying noise in animal acoustic signals. Tobias Riede (Dept. of Biology, Univ. of Utah, 257 S 1400 E, Salt Lake City, UT 84112, t.riede@utah.edu)

Variability in noisy calls or call segments is often difficult to quantify but critical in order to understand animal acoustic communication. Such components represent a large proportion of animal acoustic signals and they are behaviorally meaningful. Such elements have also collectively been labeled as nonlinear phenomena. Three approaches are reviewed. The harmonics-to-noise ratio (HNR) can be related to perceptual characteristics of an acoustic signal as well as to the mechanism of its production. The HNR seems to be a useful numeric parameter to quantify vocal variability. Two problems with the procedure will be outlined. The nonlinear measure (NLM) is designed to provide an overall estimate of the strength of nonlinearity in a signal. The rationale is that if signal noise reflects low-dimensional chaos as produced through nonlinear processes, deterministic-nonlinear modeling will produce a relatively small error component. A stochastic linear model, on the other hand, will produce a relatively large error component when applied to such signals. In a third approach, utterances can be segmented in the temporal domain into harmonic and noisy components for separate analyses. Noise-specific procedures are used to analyze nonlinear phenomena. Advantages and disadvantageous of all three procedures are exemplified with sounds from mammals and anurans.

10:40

5aAB10. Parametric and holistic approaches to analyzing primate vocal signals for acoustic markers of caller identity, emotional state, and external context. Drew Rendall and John Vokey (Dept. of Psychol., Univ. of Lethbridge, Lethbridge, AB T1K3M4, Canada)

Vocal signals can be acoustically differentiated according to a variety of intrinsic and extrinsic dimensions. Among the most commonly investigated sources of signal differentiation in primates are those that mark the identity of the caller, potential variation in the caller's underlying emotional or motivational state, and external environmental factors associated with signaling (e.g., the presense of predators or food). We analyzed a corpus of grunt vocalizations recorded from wild baboons for acoustic cues associated with these three dimensions. Grunts were collected from eight adult female baboons in two different behavioral contexts: 1. When approaching other females to handle their infants, and; 2. When embarking on an extended foraging excursion. Both contexts were further subdivided into conditions associated with low- versus high-arousal for callers. Analysis of the signals involved two different approaches: 1. A parametric analysis of specific acoustic features logically connected to either vocal-fold activity or supralaryngeal filtering (i.e., Fo and resonance properties) combined with principle components analysis (PCA) dimension reduction and discriminant analysis call classification; 2. A holistic analysis of the global spectral structure of entire signals that combined a PCA-based eigenvector decomposition and neural network classification of the calls. We compare results of the two analysis strategies.

11:00

5aAB11. Decoding the information contained in the alarm calls of Gunnison prairie dogs. Con Slobodchikoff, William Briggs, and Patricia Dennis (Dept. Biol. Sci, Northern Arizona Univ., NAU Box 5640, Flagstaff, AZ 86011, Con.Slobodchikoff@nau.edu)

Gunnison prairie dogs have been shown to produce alarm calls that incorporate information about predators [Slobodchikoff *et al.*, Anim. Behav. **42**, 713–719 (1991)], such as the species of predator and also a description of the color and general size of the predator. The alarm calls contain a series of harmonics that encode this information. The calls have been analyzed with discriminant function analysis (DFA) and classification techniques such as self-organizing neural networks. These tools have been able to show that very specific information is encoded in the calls. However, exactly how this information is encoded into the signal has proven to be elusive. We show examples of the specific information that is encoded, and present a hypothetic model of how that information is encoded into the signals.

11:20

5aAB12. Processing acoustic features of animal vocalizations using classification procedures. Sean F. Hanser, Brenda McCowan (Dept. of Vet Med. Population Health and Reproduction, Univ. of California, 1 Shields Ave., Davis, CA 95616, sfhanser@ucdavis.edu), Laurance R. Doyle (SETI Inst., Mountain View, CA 94043), and Ann E. Bowles (Hubbs-SeaWorld Res. Inst., San Diego, CA 92109)

Remote methods for classifying age, sex, group membership, or individual identification of animals that live in visually obscured environments are extremely valuable tools for field biologists, but reliable identification of individual callers still presents important challenges. Acoustic features of animal vocalizations can be processed to extract caller identification using a variety of sophisticated classification techniques, but the exact classification process can be difficult to justify rigorously and challenging to repeat on novel data. A feature extraction and classification process that is clear, simple, and repeatable would be a major benefit to wildlife studies. Classification and regression trees (CART) generate intuitive and clear processes for handling multidimensional acoustic information. Examples of CART applied to Mexican spotted owl (*Strix occidentalus lucida*) and humpback whale (*Megaptera novaeangliae*) vocalizations will be provided. These CART results will be compared to other classification techniques, particularly neural networks. CART performance was comparable, but had the advantage that it yielded explicit classifiers used to categorize vocalizations, making it easy to integrate into acoustic surveying systems. This promises to be a valuable tool for conservation and management of these and other endangered species.

11:40

5aAB13. Hidden Markov models for the analysis of animal vocalizations. Patrick J. Clemins (525 N Pollard St., Apt. 419, Arlington, VA 22203, patrick@clemins.name) and Michael T. Johnson (Marquette Univ., Milwaukee, WI 53201)

Hidden Markov models (HMMs) are commonly used to model sequential data, including time series such as vocalizations. HMMs are uniquely suited to model time series because of their ability to implicitly align samples and to be linked together to create flexible recognition patterns for vocalizations with repeating patterns, such as bird and whale songs. HMMs require less human interaction and manual tuning for successful recognition and are inherently more noise resistant than template-based techniques. Due to their statistical basis, HMM-based recognition systems can incorporate additional statistical models. One such example is a language model, which can be designed to include *a priori* knowledge about the structure of the vocalizations, such as syllable repetition patterns. HMMs have become the most popular recognition model in speech processing, and this experience has been applied to a variety of vocalization analysis tasks including individual identification, song classification, call type recognition, and physiological state determination. Classification results from complete HMM-based systems for a variety of tasks and species will be discussed.

FRIDAY MORNING, 22 MAY 2009 PAVILION EAST, 8:15 TO 11:45 A.M.

Session 5aBB

Biomedical Ultrasound/Bioresponse to Vibration, Physical Acoustics and ASA Committee on Standards: Metrology and Calibration of High Intensity Focused Ultrasound

Peter J. Kaczkowski, Chair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105-6698

Chair's Introduction—8:15

Invited Papers

8:20

5aBB1. Improved impulse response for hydrophone measurements in therapeutic ultrasound fields. Michael S. Canney (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), Vera A. Khokhlova, Oleg A. Sapozhnikov (Univ. of Washington, Seattle, WA), Yuri A. Pishchalnikov (Indiana Univ. School of Medicine, Indianapolis, IN), Adam D. Maxwell (Univ. of Michigan, Ann Arbor, MI), Michael R. Bailey, and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

The accurate measurement of pressure waveforms in high intensity focused ultrasound (HIFU) fields is complicated by the fact that many devices operate at output levels where shock waves can form in the focal region. In tissue ablation applications, the accurate measurement of the shock amplitude is important for predicting tissue heating since the absorption at the shock is proportional to the shock amplitude cubed. To accurately measure shocked pressure waveforms, not only must a hydrophone with a broad bandwidth ($>$100 MHz) be used, but the frequency response of the hydrophone must be known and used to correct the measured waveform. In this work, shocked pressure waveforms were measured using a fiber optic hydrophone and a frequency response for the hydrophone was determined by comparing measurements with numerical modeling using a KZK-type equation. The impulse response was separately determined by comparing a measured and an idealized shock pulse generated by an electromagnetic lithotripter. The frequency responses determined by the two methods were in good agreement. Calculations of heating using measured HIFU waveforms that had been deconvolved with the determined frequency response agreed well with measurements in tissue phantom. [Work supported by NIH DK43881, NSBRI SMST01601, NIH EB007643, and RFBR.]

8:40

5aBB2. Development of a high intensity focused ultrasound (HIFU) hydrophone system. Mark E. Schafer (Sonic Tech, Inc., 23 Brookline Ct., Lower Gwynedd, PA 19002, marks@sonictech.com) and James Gessert (Sonora Med. Sys., Longmont, CO 80503)

The growing use of high intensity focused Ultrasound (HIFU) has driven a need for reliable, reproducible measurements of HIFU acoustic fields. A reflective scatterer approach, incorporating several novel features for improved bandwidth, reliability, and reproducibility has been demonstrated [M. E. Schafer, J. Gessert, and W. Moore, Proc. IEEE Ultrasonics Symposium, 1739–1742 (2005)]. Several design improvements which have increase the signal-to-noise ratio of the system, and potentially reduced the cost of implementation, are now presented. For the scattering element, an artificial sapphire material is used to provide a more uniform radiating surface. The receiver is a segmented, truncated spherical structure with a 10 cm radius, made from 25 micron thick, biaxially stretched PVDF, with a Pt-Au electrode on the front surface. A specialized backing material provides the stiffness required to maintain structural stability, while at the same time providing both electrical shielding and ultrasonic absorption. This new receiver design has improved the noise performance by 8–12 dB; the new scattering sphere has reduced the scattering loss by another 14 dB, producing an effective sensitivity of −298 dB re 1 microVolt/Pa. The design trade-off still involves receiver sensitivity with effective spot size. However, the reduced cost and improved repeatability makes the overall design more robust for routine HIFU system measurements.

9:00

5aBB3. Characterization of high-intensity focused ultrasound systems by extrapolating from field measurements conducted in the quasi-linear regime. Samuel M. Howard (592 E. Weddell Dr., Sunnyvale, CA 94086, sh@ondacorp.com)

The purpose of this talk will be to discuss a method for characterizing HIFU fields where hydrophone measurements are made by reducing the electrical power input to the system until quasi-linear field propagation conditions are achieved, and until the pressure levels are safe for the hydrophone used. Radiation force measurements of total power are then used to scale up the results. This allows for reliable measurements which do not damage the hydrophone, and avoids complications from involving the many harmonics which can be generated in water which are usually not so strongly present in tissue. Advantages and disadvantages of the method will be discussed.

9:20

5aBB4. Temperature estimation and protocol assessment for monitoring high intensity focused ultrasound therapy with diagnostic ultrasound. Gavriel Speyer, Peter J. Kaczkowski, Andrew A. Brayman, and Lawrence A. Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105-6698, gavriel@u.washington.edu)

Characterizing the performance of diagnostic ultrasound (DU) for monitoring high intensity focused ultrasound (HIFU) therapy is essential for identifying effective treatment protocols. Such protocols ensure accurate therapy monitoring in media with known thermal and perfusive properties. The protocols are specified by treatment duration, transducer motion, and the DU acquisition characteristics. The HIFU transducer is constrained to move in the plane perpendicular to its axis, with the DU positioned to capture the focal plane of the HIFU transducer. Differences in RF backscattering observed between two frames, one captured before and one after treatment, are known to result from a temperature distribution generated by the heat deposited during therapy. A functional expansion for the heating is used which provides minimum variance coefficient estimates for stationary backscattering by averaging the Fisher information matrix. We show through approximation, simulation, and experiment that these functions attain the Cramer-Rao bound independent of the particular heating applied or the RF backscattering encountered. Estimator accuracy is thus determined by the material properties, including the spatial correlation, and the protocol employed. Temperature estimates accurate to well within one degree Celsius are possible, and the estimation algorithm is implemented efficiently to enable near real-time monitoring. [Work supported by NIH Grant 5R01CA109557.]

9:40

5aBB5. Ultrasound-guided high intensity focused ultrasound therapy—Safety concerns from a clinical perspective. G. R. ter Haar (Inst. of Cancer Res. and Royal Marsden NHS Trust, Downs Rd., Sutton, Surrey SM2 5PT, UK, gail.terhaar@icr.ac.uk)

The use of high intensity focused ultrasound (HIFU) in surgical applications has continued to grow at a rapid pace. Nevertheless, safety concerns limit the adoption of ultrasound guidance due to the persistent lack of real-time monitoring and quantitative measures of clinical endpoints. Imaging and temperature estimation using magnetic resonance remains the preferred approach for guidance and monitoring in much of the Western world, while ultrasound imaging has seen wide use in China. Recent reviews of clinical outcomes have raised new regulatory concerns in China and are likely to increase scrutiny elsewhere. Still lacking are near real-time measures of *in situ* parameters such as acoustic power, heating rate, or temperature, from which clinical endpoint metrics may be obtained with acceptable accuracy. Furthermore, therapy planning remains a semiquantitative process, producing highly variable results, particularly in the absence of feedback. This presentation provides an update on the clinical impact of insufficient quantification of ultrasound-guided HIFU therapy.

10:00—10:30 Break

10:30

5aBB6. *Ex vivo* monitoring of high intensity focused ultrasound treatment. Xinliang Zheng and Shahram Vaezy (Dept. of Bioengineering, Box 355061, Univ. of Washington, 1705 NE Pacific St., Seattle, WA 98195, xlzheng@u.washington.edu)

High intensity focused ultrasound (HIFU) is a therapeutic modality that can produce coagulative necrosis of tissues noninvasively. The present study investigated the feasibility of monitoring lesion formation during HIFU treatment using ultrasound backscattered signal, specifically for treatments that do not result in hyperecho in ultrasound images (i.e., thermal lesion with no obvious boiling involved). *Ex vivo* chicken breast tissue was used and a sequence of two-dimensional radiofrequency (RF) signal was collected for each exposure as a function of time. For each frame in the RF sequence, one-dimensional cross correlation coefficient was calculated between every A-mode line of RF signal and the same line 5 seconds later. A series of experiments was performed with different HIFU treatment durations. The results showed that there was a decrease in the maximum coefficient value in the focal region during HIFU exposure, and a threshold value could be determined to be associated with the time that a thermal lesion was formed. Tissue dissection was performed to confirm the presence of a thermal lesion. This RF analysis technique has shown promise in detecting location and dimension of thermally-induced lesions, as well as being capable of real-time treatment monitoring to guarantee effective and efficient HIFU therapy.

10:45

5aBB7. Impact of temperature on bubbles excited by high intensity focused ultrasound. Wayne Kreider, Michael R. Bailey (Ctr. for Industrial and Medical Ultrasound, APL, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, wkreider@u.washington.edu), Oleg A. Sapozhnikov (Moscow State Univ., Moscow 119992, Russia), and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

Bubble-enhanced heating is a current topic of interest associated with high intensity focused ultrasound (HIFU). For HIFU treatments designed to utilize acoustic radiation from bubbles as a heating mechanism, it has been reported that useful bubble activity diminishes at elevated temperatures. To better understand and quantify this behavior, a model has been implemented that couples the thermodynamic state of a strongly driven spherical bubble with thermal conditions in the surrounding liquid. This model has been validated over a range of temperature conditions against experimental data from the collapses and rebounds of millimeter-sized bubbles. With this model, the response of a micron-sized bubble was simulated under exposure to MHz/MPa HIFU excitation, while various surrounding liquid temperatures were considered. Characterizing the bubble response through the power spectral density of pressure radiated from the bubble, model calculations suggest that bubble collapses are significantly attenuated at temperatures above about 70°C. For instance, the acoustically radiated energy at 80°C is an order of magnitude less than that at 20°C. These results suggest that the efficacy of bubble-enhanced heating may be limited to temperatures below 70°C. Moreover, temperature will affect hydrophone measurements used to passively assess cavitation activity. [Work supported by NIH DK43881 and NSBRI SMST01601.]

11:00

5aBB8. Development of a tomographic cavitation sensor for quality assessment of clinical high intensity focused ultrasound systems. Stuart Faragher, Miklós Gyöngy (IBME, Univ. of Oxford, ORCB, Oxford OX3 7DQ, UK), Mark Hodnett, Adam Shaw (Natl. Physical Lab., Teddington TW11 0LW, UK), and Constantin-C. Coussios (Univ. of Oxford, ORCB, Oxford OX3 7DQ, UK)

High-intensity focused ultrasound (HIFU) fields are known to nucleate and excite inertial and noninertial cavitation in tissue and tissue-mimicking materials once a threshold negative acoustic pressure is reached. In the context of ablative HIFU treatment, inertial cavitation has been correlated with significantly enhanced rates of heating, while in histotripsy, cavitation is the very mechanism that causes tissue damage. Characterizing the extent of the cavitation region produced by clinical HIFU devices is therefore important to ensure safe, efficient, and effective treatment. A novel, multielement sensor is being developed to enable accurate axial and radial mapping of the cavitation region during HIFU exposure by passive detection and tomographic reconstruction of the broadband emissions arising from bubble collapse. Computational modeling has shown that the application of a cross-correlation algorithm to simulated received signals has the potential to localise individual sources of emissions with submillimeter accuracy. Initial experimental validation of models has been conducted using a prototype device developed in collaboration with the National Physical Laboratory. Future work will involve the refinement of the sensor design and reconstruction algorithm to improve spatiotemporal resolution, along with the development of a tissue mimicking material which matches the acoustic properties and cavitation threshold of real tissue.

11:15

5aBB9. A Schlieren system for optical visualization of ultrasonic fields. Peter J. Kaczkowski, Michael R. Bailey, Vera A. Khokhlova, and Oleg A. Sapozhnikov (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, peter@apl.washington.edu)

Ultrasonic field mapping is an essential component of transducer characterization and of beam forming verification. Such measurements are commonly performed by displacing a hydrophone over a range of points within the field; these procedures can be time-consuming. A calibrated hydrophone can provide accurate measurements of the field, subject to limitations of bandwidth and aperture of the device. A rapid qualitative 2D measurement of the spatial acoustic field can be obtained by optical means, in which the change in optical index due to the presence of acoustic pressure is imaged using a schlieren approach. This technique illuminates a transparent refracting acoustic medium using a plane collimated source and then focuses the transmitted light using a lens or mirror. In the absence of acoustic field, all of the light focuses to a small spot; acoustically induced refractive index perturbations cause some of the light to focus elsewhere. Obscuring the primary focal spot of unperturbed light with a mask permits imaging only the perturbations in the acoustic medium. We will describe a mirror-based schlieren system for imaging continuous as well as pulsed fields and with color corresponding qualitatively to the intensity of the field. [Work supported by NIH CA109556 and EB007643.]

11:30

5aBB10. Improved range-discriminating ultrasonic vibrometer. James Martin, Michael Gray, and Peter Rogers (Georgia Inst. of Technol., School of Mech. Eng., Atlanta, GA 30332-0405)

Improvements have been made to a previously reported continuous-wave ultrasonic vibrometer with range discrimination capabilities. In its current realization, the vibrometer incorporates a high-speed 16-bit arbitrary waveform generator as a signal source. The waveform generator shares a sample clock with the data acquisition system so that both the real and apparent spectral purity of each of the tones in a multitone carrier signal is preserved in the measurement. The new configuration provides increased control over the characteristics of the transmitted signal. It permits the exploitation of the full bandwidth capabilities of all of the hardware components of the system by removing a limit to the number of tones in the carrier. It also permits the relative phases of individual tones to be optimized in order to reduce the crest factor of the transmitted signal and better exploit the dynamic range of other system components. A further improvement in dynamic range has been achieved with an analog subtraction of the transmitter-receiver crosstalk, which is normally the largest received-signal component, prior to digitization. The improved vibrometer has been used to image shear wave fields in both tissue phantoms and inert tissue samples. [Work supported by ONR.]

Session 5aNS

Noise: Road Vehicle and Construction Noise—Measurement, Modeling, and Control

Kerrie G. Standlee, Chair
Daly-Standlee & Associates Inc., 4900 SW Griffith Dr., Suite 216, Beaverton, OR 97005

Contributed Papers

8:00

5aNS1. The City of Portland responds to urban gentrification through an investigation of noise sources within the boundaries of north Portland neighborhoods. Adam C. Jenkins and Julie A. Wiebusch (The Greenbusch Group, Inc., 1900 W. Nickerson St., Ste. 201, Seattle, WA 98119)

The North Portland area is a vibrant community, supporting residential populations, a wildlife reserve, and an active commerce. As Portland continues to attract new residents and residential areas expand in North Portland, the conflict between the residential desire for a peaceful environment and the commercial requirement to efficiently conduct business become more critical. The sustainability of the North Portland residential communites relies on planning decisions based on a well-defined understanding of the sound characteristics of the community. The City of Portland chose to study North Portland to better understand noise within its boundaries. The resulting noise mapping will serve as a tool for the City to use in future planning.

8:15

5aNS2. Noise mapping in Portland Oregon: The need to develop an approach to accurately depict the health impacts of short duration automotive racing noises emanating from a variety of motorized race events at the Portland International Racetrack facility. Paul van Orden (City of Portland Oregon, 1900 SW 4th Ave., Portland, OR 97201, pvanorden@ci.portland.or.us)

The City of Portland Oregon is in the final phase of completing its efforts to map one quadrant of the city in conjunction with the Greenbusch Group of Seattle Washington. The selection of this section of the city was identified for its notable concentration of noise sources ranging from the normal transportation related road, railway, and international airport sounds, to the unique and constantly varying tonal quality of automobile and motorcycle race groups at an international automotive raceway facility. The challenge to identify an appropriate method or standard to accurately depict the impact of these short term track races and drag races offers a unique opportunity to expand on the value of the noise mapping approach for long term urban planning efforts.

8:30

5aNS3. Prediction of noise from Portland International Raceway using environmental noise computer modeling techniques. Angus M. Deuchars (Arup Acoust., 560 Mission St., Ste. 700, San Francisco, CA 94105, angus.deuchars@arup.com)

An environmental noise computer model of Portland International Raceway has been constructed to evaluate noise mitigation at local noise sensitive receptors from making site wide changes to raceway infrastructure. There is little precedent of environmental noise computer modeling of raceways, and this paper presents the methodologies that were used to develop the computer model. Field measurements were made at the racetrack according to European Union Harmonoise methods to determine sound power levels of race vehicles; noise predictions were made using ISO 9613 pt 2 prediction model implemented using SOUNDPLAN software. Results from field measurements and predictions are presented to demonstrate accuracy of the computer model. Additionally, the innovative steps used to construct the computer model using sound plan are also presented.

8:45

5aNS4. A revised standard method for ground impedance measurement. Keith Attenborough and Shahram Taherzadeh (Dept. of Design, Development, Environment & Mater., The Open Univ., Milton Keynes MK7 6AA, UK, k.attenborough@open.ac.uk)

ANSI S1.18 (1998) for ground impedance measurement has been revised to combine the previously existing template method with direct deduction of impedance from data for the complex level difference between vertically separated microphones. It has been found that, despite the improved sensitivity at low frequencies that should be achievable through use of a longer range, the combined approach ceases to give reliable information for horizontal source-receiver separations of more than 1 m. The revised standard contains four worked examples: three grass-covered ground surfaces and a gravel road. The first example has been used to show that although use of consecutive measurements with a single microphone produces data that is usable with the template method, the phase information is not sufficiently robust to allow direct impedance deduction. In the three grassland examples, the widely used single parameter model is found to overpredict the real and imaginary parts of ground impedance at lower frequencies. The fourth example shows that the geometries prescribed in the standard are not appropriate for determining impedance spectra of relatively hard surfaces but are sufficient to establish that the surface is acoustically hard.

9:00

5aNS5. Sampling strategies for incorporating uncertainty and turbulent scattering into outdoor sound propagation calculations. Matthew S. Lewis, D. Keith Wilson (U.S. Engineer Res. and Development Ctr., 72 Lyme Rd., Hanover, NH 03755, D.Keith.Wilson@usace.army.mil), and Chris L. Pettit (U.S. Naval Acad., Annapolis, MD 21402)

The accuracy of mean sound-level estimates derived from different sampling methods is studied. The methods considered are Monte Carlo sampling (MCS), Latin Hypercube sampling (LHS), and Latinized Centroidal Voronoi Tessellation sampling (LCVTS). The goal is to determine which sampling method converges most rapidly to the actual mean sound level. The approach involves a model acoustic atmosphere, based on surface-layer similarity theory and a relaxation model for the ground, with sound fields calculated numerically with a parabolic equation solver. The range of consequences of epistemic uncertainty due to imperfect knowledge of the atmospheric and ground variables is examined through a simplified probabilistic model. Random (aleatory) uncertainty due to turbulent scattering also is considered. The samples are drawn randomly within the domain of uncertainty of the environmental variables. When only epistemic uncertainty is present owing to imprecise knowledge of the environmental variables, LCVTS is found to converge to an accurate estimate with the fewest number of samples, followed by LHS and then MCS. When random turbulent scattering is present, however, the sampling method has little effect on convergence.

9:15

5aNS6. Construction noise action area model. Larry J. Magnoni (Wash. St. Dept. of Trans., Env. Svs., Acoust., Air Quality, Energy Sec., 15700 Dayton Ave. N., PO Box 330310, Shoreline, WA 98133, magnonil@wsdot.wa.gov)

A model developed by the Washington State Department of Transportation to be used by biologists to determine the action area for highway construction projects is presented. The model conservatively estimates the distance at which construction noise would match ambient background sound levels. The model propagates the combined maximum sound level (Lmax) for project construction equipment spherically over the soft and hard site characteristics to a point where it would converge with the ambient hourly equivalent (Leq) sound level. The ambient Leq sound level can be input into the model from actual measurements, documented reference sources, or from a table based on population density [Federal Transit Authority (FTA) Noise Assessment Guidance (2006)]. The model then combines the ambient sound level with the traffic sound level based on the national mean emission curves. The input for traffic volume, vehicle type, and speed is propagated cylindrically from the source until the traffic source drops to the ambient level. A macro, that applies a general atmospheric and molecular absorption adjustment to the calculated distances, produces the action area circumference both graphically and numerically as it relates to each of the sound sources. [Work supported by the Federal Highway Administration (FHWA) (1978).]

9:30

5aNS7. Analysis of pile driver exhaust and impact noise. Edward Zechmann and Charles Hayden (Robert A. Taft Bldg., 4676 Columbia Pkwy., C-27, Cincinnati, OH 45226)

In July 2008, NIOSH accomplished a preliminary noise survey focused on gathering impulsive noise data from pile drivers. The purpose of the noise survey was to better understand the noise characteristics of pile drivers. NIOSH gathered time record data of H-beam piles being driven into hard rock. This data was used to characterize the frequency content of the major impulsive events (impact and exhaust cycles). Knowing the frequency content is necessary to apply existing or develop new noise controls, since many noise controls are frequency dependent. Gathering the frequency content of a pile driver's impulsive noise signal is typically difficult using off-the-shelf measurement and analysis equipment and software. Reliable time record data is necessary to confidently extract the frequency content from the signal. NIOSH has developed an Impulsive Noise Meter laptop program to properly gather and analyze these impulsive signals. By analyzing the third octave band peak levels of each impact and exhaust cycle of the pile driver operation, the frequency dependence was determined. Results of this analysis showed the frequency content varied as the pile was driven further into the ground. This implies that broad spectrum noise controls are necessary to control those noise emissions.

9:45

5aNS8. Are hybrid cars too quiet? Ryan L. Robart and Lawrence D. Rosenblum (Dept. of Psych., Univ. of California, Riverside, 900 University Ave., Riverside, CA 92521)

The increase in availability of alternative fuel vehicles has elicited concerns for pedestrians who might not hear the approach of these quieter cars. Three experiments tested the relative audibility of hybrid vehicles (in their electric mode) and internal combustion engine (ICE) cars. Binaural recordings were made of the cars approaching from either the right or left, at 5 mph. Subjects were asked to listen to these recordings over headphones and press one of two buttons indicating from which direction the car approached. Subjects' accuracies and reaction times were measured. The first experiment revealed that (sighted) subjects were able to determine the approach direction of the ICE cars substantially sooner than the hybrid cars. A second experiment added the natural background sounds of idling engines to the stimuli. The addition of background sound disproportionately hindered perception of the hybrid cars, so that they could not be localized until very close. A final experiment testing both sets of stimuli with blind subjects revealed the same pattern of results. Implications of these results for pedestrian safety will be discussed. [Work supported by a grant from the National Federation for the Blind.]

10:00—10:15 Break

10:15

5aNS9. Impulse noise reduction for hearing protectors. William Murphy (Natl. Inst. for Occupational Safety and Health, Div. Appllied Res. and Technol., Hearing Loss Prevention Team, 4676 Columbia Pkwy., MS C-27, Cincinnati, OH 45226-1998, wjm4@cdc.gov)

In 2009, the United States Environmental Protection Agency will propose a revision to the federal regulation for the labeling of hearing protection devices, 40 CFR 211 Subpart B. One of the new features of the proposed rule was the measurement of an impulse noise reduction rating for hearing protection devices. Measurement of impulsive sounds is challenging technically from an acoustics perspective. This paper will report on the performance of an acoustic shock tube used to generate impulses between 140 and 170 dB peak sound pressure level. The calibration methods for the microphones will be discussed and the measurements for a variety of hearing protectors will be presented. Typical earmuffs are capable of impulsive noise reduction ratings of between 20 and 35 dB. Earplugs provide similar range of performance. Combinations of earmuff and earplug have yielded impulse peak reductions of more than 50 dB. The reduction of the impulse peak level should provide a means to predict exposure at the ear when a hearing protector is worn in an impulsive noise environment.

10:30

5aNS10. Noise barriers based on recycled materials. Jose Sanchez-Dehesa, Victor Garcia-Chocano, Daniel Torrent, Francisco Cervera, and Suitberto Cabrera (Wave Phenomena Group, Polytechnic Univ. of Valencia, Camino de vera s.n., ES-46022 Valencia, Spain, jsdehesa@upvnet.upv.es)

Two-dimensional sonic crystals consisting of arrays of cylinders fabricated with recycled materials (rubber crumb) are proposed to reduce efficiently the noise. The attenuation by these barriers is produced by a combination of sound absorption by the rubber crumb and Bragg reflection by the ordered cylinders. Experiments performed in an anechoic chamber support the predictions by the multiple scattering theory. An optimization method is also reported to develop sound barriers for traffic noise. [Work supported by Spanish MICIIN.]

10:45

5aNS11. Quenching of bandgaps by flow noise. Jose Sanchez-Dehesa (Wave Phenomena Group, Polythecnic Univ. of Valencia, Camino de vera s.n., Spain, ES-46022, jsdehesa@upvnet.upv.es), Tamer Elnady, Adel Elsabbagh, Wael Akl, Osama Mohamady (Ain Shams Univ., Cairo, Egypt), Victor Garcia-Chocano, Daniel Torrent, and Francisco Cervera (Polythecnic Univ. of Valencia, ES-46022, Spain)

We report an experimental study of acoustic effects produced by wind impinging on noise barriers based on two-dimensional sonic crystals with square symmetry. We found that the attenuation strength of sonic-crystal-bandgaps decreases for increasing values of low speed. A quenching of the acoustic bandgap appears at a certain speed value that depends of the barrier filling ratio. For increasing values of low speed, the data indicate that the barrier becomes in a sound source because of its interaction with the wind. We conclude that flow noise has paramount importance in designing acoustic barriers based on sonic crystals. [Work supported by Spanish AECI.]

11:00

5aNS12. Leakage identification for a vehicle firewall sound transmission by beamforming technique. Yuri Ribeiro (GM do Brasil, Indaiatuba, Sao Paulo, Brazil), William Fonseca, and Gerges Samir (Fed. Univ. of Santa Catarina, Florianopolis, SC, Brazil)

The acoustic phased array is a powerful tool for identification and quantification of noise sources on complex systems. Acoustic sources generate waves which propagate through space to the N microphones of the array. The acoustic pressures are sampled and converted to time series data and processed using either time domain or frequency domain delayed and averaged. The output of signal processing is a colored image with relative sound pressure levels distribution. In this paper application of beamforming technique is used to identify leakage areas in a vehicle firewall. Two acoustic rooms are used and the firewall was installed in the apertures inbetween with careful control of noise flank paths. Source room was excited with white noise and a phased array microphone was used to measure the sound field in the anechoic receiving room. The firewall was evaluated both with all holes sealed and with the regular pass-through components installed. Two different acoustic insulators were also tested on the firewall. The results show clearly the areas of leakage with relative noise levels at each test configuration.

FRIDAY MORNING, 22 MAY 2009 BROADWAY III/IV, 8:30 TO 10:00 A.M.

Session 5aPA

Physical Acoustics: General Physical Acoustics II

Kai Ming Li, Chair
School of Mechanical Engineering, Ray W. Herrick Labs., Purdue Univ., West Lafayette, IN 47906

Contributed Papers

8:30

5aPA1. Using sound radiation to characterize binary collisions of polypropylene balls. J. Riner and A. Petculescu (Dept. of Phys., Univ. of Louisiana, P.O. Box 44210, Lafayette, LA 70504)

An acoustic method to check the validity of Hertz's law and characterize the source characteristics of binary collisions is proposed. The technique uses a goniometer equipped with a microphone to detect the sound produced by collisions between two solid polypropylene balls. The collisions are the result of a projectile ball falling from rest and striking a rigidly mounted target ball. The acoustic field in both the horizontal and the vertical plane is sampled and the maximum and minimum pressures, as well as frequency content, are studied. The power spectrum has a dominant peak at approximately 10 kHz. Radiation in the horizontal plane is found to be mostly symmetric about the collision point while there is an observed asymmetry in the vertical plane, which can be used to infer the characteristics of the source of audible sound. The acoustic signals are likely generated by the relative accelerations of the two balls and not by the normal vibrational modes, which were measured to start at about 73 kHz. [Work funded by the Louisiana Board of Regents.]

8:45

5aPA2. Pulse propagation in a hyperlattice. Joseph Dickey (Adv. Technol. Lab., The Johns Hopkins Univ., 810 Wyman Park Dr., Baltimore, MD 21211)

The classical dynamics and pulse propagation are presented for a series of latticelike structures whose spatial dimensionality ranges from one to four: four representing a hyperlattice. The lattices are connected 1-D wave bearing systems of varying lengths and can illuminate some aspects of higher dimension structures. Short pulses are launched at an arbitrary point, reverberate throughout the entire structure, and detected at another point. Some aspects of increasing dimensionality are illustrated with particular emphasis on the transition from three to four spatial dimensions. In a hypothetical 4-D world where only three are observable, the classical conservation laws and causality do not hold. The lack of causality is illustrated at each step in dimensionality by showing the "unexpected" pulse returns from the next higher dimension.

9:00

5aPA3. Scattering of compressional waves by a cavity in an elastic solid: Geometric interpretation and beam illumination. Likun Zhang and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814, zhanglikun@wsu.edu)

Aspects of the scattering of short-wavelength compressional waves by an empty spherical cavity in an elastic solid may be predicted geometrically from the elastic wave reflection coefficient for a plane surface and the curvature of the reflecting surface. When the outgoing wave is a shear wave, it is necessary to include the modified ray-tube parameters associated with mode conversion. Related geometrical considerations are also helpful for interpreting the scattering of a compressional wave Bessel beam by an empty spherical cavity centered on the beam in an elastic solid. The exact solution in that case follows from a generalization of the scattering analysis for acoustic Bessel beams [P. L. Marston, J. Acoust. Soc. Am. **121**, 753–758 (2007)]. A modified analysis gives the directional dependence of the shear waves resulting from mode conversion for a cavity on the axis of a compressional wave Bessel beam. For both plane wave and Bessel beam illumination, the scattering pattern can depend strongly on the value of the Poisson ratio of the solid.

9:15

5aPA4. Measurement of acoustic wave frequency by Bragg light diffraction. R. Farkhad Akhmedzhanov (Dept. of Phys., Navoi State Mining Inst., 27a Yuzhnaya St., Navoi 210100, Uzbekistan, farkhad2@yahoo.com)

The new technology is offered for measurement of acoustic wave frequency. The method is based on using the Bragg light diffraction on the hypersonic transversal acoustic wave in a gyrotropic crystal. As is well known, the intensity of diffracted light in this case is strongly dependent from the quantity of specific rotation of polarization vector. As a result, using the standard calibration of light intensity it is possible to measure the changes of acoustic wave frequency. Measurements of the diffracted light intensity have been carried out in several points along the direction of the acoustic wave propagation. The applied sample of LiNbO3 was oriented along the crystallographic axis of the third order with the accuracy of 10′. The plane-polarized transverse acoustic waves were excited in the frequency range from 900 MHz to 1.2 GHz. The results obtained at the average frequency 1.0 GHz were shown, that the intensity of diffracted light intensity is sensitive parameter to the change of the transverse acoustic wave frequency. The dependence of the diffracted light intensity from the quantity of specific rotation frequency was used for determination of the acoustic wave frequency by the appropriate model equation accurate within 0.1%.

9:30

5aPA5. Acoustic cavities with radial sonic crystals. Daniel Torrent and José Sánchez-Dehesa (Wave Phenomena Group, Polytechnic Univ. of Valencia, Camino de Vera s.n. ES-46022, datorma1@upvnet.upv.es)

Radial sonic crystals in two and three dimensions are introduced. These materials have the property of leaving the acoustic wave equation invariant under radial displacements, and then they present a "radial band structure" similar to that of one dimensional sonic crystals. Finite shells of radial sonic crystals can be used as resonant cavities. The radiation pattern of these cavities will be reported. For example, it will be shown that a cavity can be used as a dipolar source by placing a monopole inside the cavity. Finally, physical realization of radial sonic crystals is discussed in two and three dimensions [Work supported by Spanish MICIIN and Consolider Program.]

9:45

5aPA6. Acoustic cloaking in two and three dimensions. Daniel Torrent and José Sánchez-Dehesa (Wave Phenomena Group, Polytechnic Univ. of Valencia, Camino de Vera s.n. ES-46022, datorma1@upvnet.upv.es)

The cloaking theory is analyzed in two and three dimensions. It is shown that cloaks require materials with some special properties not known in nature, mainly, they need to be fluid-like materials with anisotropic, inhomogeneous, and locally divergent acoustic properties. The anisotropy problem is solved with the help of the so-called "metamaterials", the inhomogeneous behavior is implemented with stratified scatterers and the divergence of the parameters is avoided with the concept of "imperfect cloaks", which transforms the scattering cross section of an object of radius R in that of an object of radius $R' \ll R$. [Work supported by the Spanish MICIIN and the Consolider Program.]

FRIDAY MORNING, 22 MAY 2009 PAVILION WEST, 8:30 TO 11:05 A.M.

Session 5aUW

Underwater Acoustics: Sediment Acoustics

Kevin L. Williams, Chair
Applied Physics Lab., Univ. of Washington, Seattle, WA 98105

Chair's Introduction—8:30

Contributed Papers

8:35

5aUW1. The effects of scattering from heterogeneities in porosity during sound propagation through sand sediments. Brian T. Hefner, Darrell R. Jackson (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698), and Joseph Calantoni (Naval Res. Lab., Stennis Space Ctr., MS 39529)

Sand sediments are inherently heterogeneous due to the random packing of the grains. For sound propagation through fluid-saturated sediments, these heterogeneities may lead to scattering from the coherent fast compressional wave into incoherent slow compressional waves or shear waves. This loss of energy from the fast compressional wave may account for the increase in high-frequency attenuation above that predicted by Biot theory. In a previous talk, we presented preliminary results of applying perturbation theory to Biot theory in order to model scattering from heterogeneities in the porosity [Hefner *et al.*, J. Acoust. Soc. Am. **120**, 3098 (2006)]. This theory has since been refined to properly account for scattering into both the slow compressional wave and the shear wave. In order to apply this theory to a given sand sediment, knowledge of the covariance function for the spatial variations in the porosity is required. Results of the theory will be presented for several different analytic covariance functions, as well as for covariance functions measured in simulated and real unconsolidated granular materials.

8:50

5aUW2. Sand acoustics: The effective density fluid model, Pierce/Carey expressions, and inferences for porous media modeling. Kevin Williams (Appl. Phys. Lab., College of Ocean and Fishery Sci., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, williams@apl.washington.edu)

Recently, Pierce and Carey [JASA-EL **124**, EL308–EL312 (2008)], presented a low-frequency analysis of sound propagation in sand/silty sediments. Here, equivalent expressions are presented using a low frequency expansion of an unconsolidated version of Biot porous medium theory (i.e., the effective density fluid model (EDFM)). The resulting expression for attenuation allows identfication of the nondimensional parameter in the Pierce /Carey result in terms of physical parameters. One of the important aspects of the Pierce/Carey derivation is that they show how to account for the inertial effects due to the relative grain/fluid density. The agreement between the Pierce/Carrey and EDFM low-frequency attenuation expressions motivates further frequency and relative grain/fluid density analyses of attenuation and sound speed. The general inference of these analyses is that any model that ignores inertial effects due to differences in grain and fluid density is missing a key piece of physics that makes it incomplete. Wood's equation is one such model and since the Buckingham model uses Wood's equation, it is another example. The point being that the use of the mass density to handle inertia in a porous medium is a low-frequency approximation. [Work supported by the Office of Naval Research.]

9:05

5aUW3. Acoustic measurement of fines. Wayne Carpenter (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, wocarpen@olemiss.edu), Daniel Wren, Roger Kuhnle (USDA-ARS-NSL, Oxford, MS 38655), and James Chambers (Univ. of Mississippi, University, MS 38677)

The measurement of sediment particles $<$100 microns in diameter suspended in water has received little attention; however, particles of this size can account for a large fraction of the material transported in streams and rivers. In order to advance the current state of knowledge, the use of acoustic techniques as a surrogate measurement for the concentration of fine sediment particles in water is investigated. Both backscatter and attenuation will be evaluated as potential measurement techniques using a laboratory tank. Based on experience with the use of 1–5 MHz frequencies for backscatter measurements of particles in the 100–1000 micron size range, initial work will focus on the use of a frequency of approximately 20 MHz. The attenuation and backscatter due to a range of concentrations of bentonite and kaolinite will be reported.

9:20

5aUW4. High-frequency acoustic properties of water-saturated sediments: Laboratory study. Jean-Pierre Sessarego, Rgine Guillermin (CNRS/LMA, 31 Chemin Joseph Aiguier, 13402 Marseille, France), and Anatoly N Ivakin (Univ. of Washington, Seattle, WA 98105)

This study was performed in controlled laboratory conditions using several broadband transducers to cover a wide frequency range from 200 kHz to 3MHz. Various types of sediments were used: natural medium and coarse sand, and glass beads with corresponding mean sizes. Sound speed and attenuation were measured and compared with predictions of Biot theory. The

theory failed to predict both high attenuation and negative dispersion of sound speed observed at very high frequencies (above 500 kHz for medium sands), which can be assumed an effect of strong grain scattering. All the measurements were performed at different times to ensure the repeatability of the results. A comparison of results obtained for natural sands with those for artificial sediment (glass beads) of the same mean size, allowed observing the influence of the grain shape on the reflected and backscattered signals. Also, the influence of the transducer directivity diagram on the sediment reflectivity was observed. This effect is discussed and shown to be essential in the case of strong incoherent scattering observed for coarse sediments. Finally, the time-frequency contents of the reflected signal were analyzed. It was found to be very sensitive to the granular structure of the sediment. [Work supported by CNRS and ONR.]

9:35

5aUW5. Direct numerical simulation of the acoustic virtual mass of granular media. Joseph Calantoni (Marine Geosciences Div. Naval Res. Lab., Stennis Space Ctr., MS 39529, joe.calantoni@nrlssc.navy.mil) and B. Todd Hefner (Univ. of Washington, Seattle, WA 98105-6698)

The theory of acoustic virtual mass in granular media hypothesizes that during sound propagation, energy is transferred out of the direction of propagation and into rotational and transverse degrees of freedom that act as an additional virtual mass, thus decreasing the sound speed. We use a discrete element model that explicitly simulates the motions and interactions of every particle in a finite volume to examine the theory of acoustic virtual mass. Simulations are fully three-dimensional where the physical properties, such as size, density, and shape (typically spherical), of every grain may be uniquely specified. Normal and tangential forces at grain-grain contact points are modeled with springs and friction, respectively. Depending on the material of interest (e.g., quartz grains), several normal contact force laws are available including linear elastic, linear elastoplastic, and Hertzian. Likewise, the tangential contact force is typically modeled with a stick-slip behavior using a linear spring and friction, but other variations are available. Sound speed and attenuation are calculated by propagation of pulses through the simulated materials and these results are compared to the predictions of theory with and without the virtual mass term.

9:50

5aUW6. Synthesizing the shape of sand grains. David R. Barclay and Michael J. Buckingham (Scripps Inst. of Oceanogr., Marine Phys. Lab., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238, dbarclay@mpl.ucsd.edu)

Digitized outlines of sand grains from a dozen locations have been acquired using an optical microscope. A Fourier decomposition of the outline is calculated providing a spectral description of the grain's shape. By averaging over several hundred grains, the normalized power spectrum of each sand sample is returned. The desert sands, beach sands, and marine sediments measured all exhibit the same inverse-power-law dependence on the harmonic number, n, varying as $n^{-10/3}$ for $2 \leqslant n \leqslant 20$. This "universal" spectrum provides the basis of a numerical technique for synthesizing the irregular outline of a sand grain: the outline is represented as a random pulse train in which identically shaped microasperities, with normally distributed amplitudes, are randomly superimposed on the perimeter of a circle. Carson's theorem links the power spectrum of an individual microasperity to that of the real grains and constrains the synthetic outlines to have the identical statistical properties as the outlines of the measured samples. Visually, it is difficult to distinguish between the synthesized and real outlines. This numerical technique for synthesizing irregular outlines of grains has potential for investigating the random-packing of realistically rough particles through computer simulation. [Work supported by the Office of Naval Research.]

10:05—10:20 Break

10:20

5aUW7. Uncertainty estimation of sound attenuation in marine sediments at low frequencies. Yong-Min Jiang and N. Ross Chapman (School of Earth and Ocean Sci. , Univ. of Victoria, P.O. Box 3055 Victoria, BC, V8W 3P6 Canada, minj@uvic.ca)

Marine sediment attenuation at low frequencies (under 5 kHz) is generally difficult to be directly measured by *in situ* probes embedded in the sediment, partly due to the very short propagation distances. An alternate experimental technique is to use single bottom bounce signals received by a vertical line array. The frequency dependence of the sediment attenuation is first obtained by comparing the amplitude differences of the sea floor reflection and the sub bottom layer reflection at different frequencies. The absolute attenuation is then obtained by using the previously estimated sound speed and layer thickness. Inherently there is uncertainty introduced in each stage of the attenuation estimation procedure. To evaluate the uncertainty of the attenuation estimate, the standard deviation of the signal fluctuation is mapped to the intermediate result first, and then Bayesian inversion results of the sound speed and the layer thickness are included in the final attenuation estimates. This uncertainty analysis is demonstrated by the estimation of the sediment attenuation from the low frequency chirp data collected in a variable water column environment in the Shallow Water 06 experiment. [Work supported by ONR Ocean Acoustics.]

10:35

5aUW8. On the use of acoustic particle motion in geoacoustic inversion. Steven E. Crocker (Naval Undersea Warfare Ctr., 1176 Howell St. Newport, RI 02841), James H. Miller (Univ. of Rhode Island, Narragansett, RI 02882), Paul C. Hines, and John C. Osler (Defence R&D Canada Atlantic, Dartmouth, NS B2Y 3Z7, Canada)

Geoacoustic inversion estimates sediment properties based on one or more parameters of an observed acoustic field via inverse mathematical methods. The observed acoustic parameter is usually derived from the acoustic pressure measured at one or more locations. Recent advances in acoustic sensor technology have enabled the simultaneous measurement of the acoustic pressure and particle motion in three dimensions. The additional information provided by such acoustic vector sensors can be used to improve the performance of existing and novel geoacoustic inversion techniques. Current research seeks to use the additional information that is provided by the acoustic vector sensor to pose new inverse problems for the estimation of seabed sediment properties. In particular, data collected during the Sediment Acoustics Experiment 2004 (SAX04) included acoustic pressure and particle acceleration from a small number of acoustic vector sensors arranged in a vertical line, spanning the water-sediment interface. A variety of waveforms were transmitted into the seabed to measure the direct, reflected, and transmitted waves. Experiment details, forward modeling, inversion methods, and results are discussed. [Work supported by ONR.]

10:50

5aUW9. Study of interface wave generation and shear wave velocity estimation. Hefeng Dong (Dept. of Elect. and Telecom., Norwegian Univ. of Sci. and Technol., NTNU, No-7491 Trondheim, Norway and School of Earth and Ocean Sci., Univ. of Victoria, BC, V8W 3P6 Canada) and Ross Chapman (Univ. of Victoria, BC, V8W 3P6 Canada)

Estimation of the shear wave velocity profile versus depth and shear wave attenuation in the upper sediment layers has been an important research topic in underwater acoustics. The shear wave velocity profile is insensitive to hydrophone data measured in water column, while dispersion of the interface wave propagating along the water and sea bottom boundary (the Scholte wave) is closely related to shear wave velocity variation with depth in the upper sediment layers. However, due to the exponential decay of the amplitude of the interface waves, there may be no interface waves generated and recorded in a conventional underwater experiment configuration where the sources and receivers (vertical and/or horizontal hydrophone array) are located in the water column with distance (measured by wavelength of the interface wave) away from water bottom interface. In this paper the conditions in which interface waves can be generated, recorded, and visualized are studied by numerical experiments in the underwater environment. The conditions include source strength, frequency band, source–receiver configuration, and signal-to-noise ratio, etc. Examples are given for estimating of shear wave velocity profile with depth by inverting dispersion curves of recorded interface waves in real data. [Work supported by NFR under Contract No. 186923/I30.]

Session 4pSW

Speech Workshop: Cross Language Speech Perception and Linguistic Experience: Poster Session A

Linda Polka, Chair
School of Communication Sciences and Disorders, McGill Univ., Montreal, QC H3G 1A8, Canada

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 even-numbered papers will be at their posters from 2:30 p.m. to 4:00 p.m.

4pSW1. Acoustic correlates of Tone 3 and Tone 4 in Mandarin. Chia-Hsin Yeh (1627 Spartan Village Apt. L, East Lansing, MI 48823)

Perceptual similarity between Tone 2 and Tone 3 in Mandarin was widely discussed in previous studies [Moore and Jongman (1997), Huang (2004), Bent (2005)]. Other tonal contrasts are hardly addressed. However, recent findings of Mandarin tones show that Tone 3 and Tone 4 are confusing in terms of descent slope. A big difference between previous studies and current ones is kinds of Tone 3 stimuli: previous studies used isolated Tone 3, namely, 214, and current studies used derived Tone 3, namely, 21, which is coarticulated with other tones. Spectrographic analysis was conducted. Descent slope and height of starting pitch were found critical between Tone 4 and derived Tone 3. Then three continua with ten instances for each were synthesized in respect of three different factors, descent slope, height of starting pitch, and height of ending pitch, which makes 30 stimuli total in three experiments, an AXB discrimination test, an identification test, and a lexical decision test. Results show that starting pitch height is a salient acoustic factor for 30 Mandarin natives, although descent slope was predicted more important. In addition, lexicon has an influence on ambiguous instances between endpoints of Tone 3 and Tone 4. Current findings may make contributions to understanding of perception, articulation, and lexicon during speech processing.

4pSW2. Intelligibility of foreign-accented speech in noise for younger and older adults. Sarah Hargus Ferguson (Dept. of Speech-Lang.-Hearing: Sci. and Disord., Univ. of Kansas, Dole Ctr., 1000 Sunnyside Ave., Rm. 3001, Lawrence, KS 66045, safergus@ku.edu), Kyung Ae Keum, Allard Jongman, and Joan A. Sereno (Univ. of Kansas, Lawrence, KS 66045)

In a previous experiment [Ferguson *et al.*, J. Acoust. Soc. Am. **118**, 1932 (2005)], young listeners with normal hearing, older adults with essentially normal hearing, and older adults with hearing loss were found to be similarly affected by the presence of a foreign accent on a word identification task in various listening conditions. This result stands in sharp contrast with an extensive literature demonstrating that the negative effect of speech distortions such as noise, reverberation, and time compression is disproportionately great for older versus younger adults. The conclusions of the previous study, however, were tempered by the presence of floor effects in the data for the older adults with hearing loss identifying accented speech in noise (12-talker babble at a signal-to-babble ratio of +3 dB). The present experiment extended the earlier one by presenting the same monosyllabic word materials to new groups of young normal, elderly normal, and elderly hearing-impaired listeners in three new signal-to-babble ratios: +6 dB, +9 dB, and +12 dB. The results will be discussed in terms of their implications for older adults interacting with foreign-accented speakers in healthcare settings. [Project supported by University of Kansas General Research Fund.]

4pSW3. The aquisition of sonorant voicing. Jagoda Sieczkowska (Inst. of Nat. Lang. Processing, Univ. of Stuttgart, Azenbergstr. 12 D-70174 Stuttgart, Germany, jagoda.sieczkowska@ims.uni-stuttgart.de)

In the view of Exemplar Theory phonetic variants of words are stored by speakers in their memory as a set of exemplars organized into clusters and grow stronger or decay over time depending on language experience (Bybee, 2002). In Polish and French sonorant consonants in clusters are devoiced word finally, which does not occur in German and English because of phonotactical constraints. A study based on automatic speech alignment and voicing profile extraction from Polish, French, German, and English speech corpuses aims at defining voicing sonorant temporal structure and voicing profile acquisition. The hypothesis is that Polish and French native speakers transfer their L1 exemplars during L2 acquisition of English and German and thus devoice sonorant. After extracting voicing profiles from recordings of native Polish and French speakers speaking English and German as an L2/L3, comparison of voicing profiles of English and German native speakers will be made in order to define the exemplar transfer.

4pSW4. Perceptual adaptation to sinewave-vocoded speech across languages. Tessa Bent (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405, tbent@indiana.edu), Jeremy L. Loebach (Macalester College, Saint Paul, MN 55105), Lawrence Phillips, and David B. Pisoni (Indiana Univ., Bloomington, IN 47405)

Listeners rapidly adapt to many forms of degraded speech. What level of information drives this adaptation (e.g., acoustic, phonetic, lexical, syntactic), however, remains an open question. In the present study, three groups of listeners were passively exposed to sinewave-vocoded speech in one of three languages (English, German, or Mandarin) to manipulate the level(s) of information shared between the training languages and testing language (English). Two additional groups were also included to control for procedural learning effects. One control group received no training, while the other control group was trained with spectrally rotated English materials. After training, all listeners transcribed eight-channel sinewave-vocoded English sentences. The results demonstrated that listeners exposed to German materials performed equivalently to listeners exposed to English materials. However, listeners exposed to Mandarin materials showed an intermediate level of performance; their scores were not significantly different from the German or English groups but were also not significantly different from the two control groups. These results suggest that lexical access is not necessary for perceptual adaptation to degraded speech, but rather similar phonetic structure between the training and testing languages underlies adaptation to degraded speech. [Work supported by NIH-NIDCD T32 DC-00012 and R01 DC-000111.]

4pSW5. Assessment of perceptual assimilation before and after training. James D. Harnsberger (Dept. of Commun. Sci. and Disord., Univ. of Florida, Gainesville, FL 32611, jharns@ufl.edu)

Many non-native speech sounds are challenging to perceive and, ultimately, to acquire. The development of a model of cross-language speech perception and learning has been hampered by, among other issues, the high variability commonly observed in the perceptual assimilation of non-native contrasts due to phonetic context, talker, and a variety of indexical properties of speech. This variability may reflect persistent patterns in perception that influence learning, or it may represent only an early sensitivity to low-level phonetic detail that is attenuated as experience increases. To examine this issue, the perceptual assimilation by native speakers of American English of a variety of voicing and place contrasts from Hindi was assessed be-

fore and after training in a paired-associate word learning task. Training utilized tokens from four of the six talkers used in perceptual assimilation tasks. The purpose of the study was to determine whether or not assimilation patterns could be greatly modified and/or simplified over the course of word learning in the laboratory. If limited training resulted in large changes in perceptual assimilations, then models of cross-language speech perception and learning will need to rely on assimilation data elicited via procedures that result in stable and repeatable patterns.

4pSW6. Production of disyllabic Mandarin tones by children. Puisan Wong (Dept. of Speech Commun. Arts and Sci., Brooklyn College, CUNY, 2900 Bedford Ave., Brooklyn, NY 11210, pswRes.@gmail.com) and Winifred Strange (CUNY Grad Ctr., New York, NY 10016)

Two- to six-year-old children's productions of Mandarin lexical tones in familiar disyllabic words were examined to determine the time course of tone development and the effect of complexity of fundamental frequency contours (f0) on rate of acquisition. Disyllabic tone (DT) productions of 44 children and 12 of their mothers were elicited in a picture naming task and were low-pass filtered to eliminate lexical information. Judges categorized the DTs based on the filtered stimuli without lexical support. Adults' and children's DTs were categorized with 96% and 65% accuracy, respectively. Tone accuracy increased while intersubject variability decreased with age. Children as old as 6 years did not attain adultlike accuracy. Different DT combinations developed at different rates. DTs with more complex f0 contours were more difficult for children. Substitution patterns and acoustic analysis revealed that when producing tones with large transitions at the syllable boundary, children tended to modify the f0 in the first syllable to reduce the f0 shift at the boundary of the first and second syllable, resulting in more errors in the tones in the first syllable than in the second syllable. The results suggest physiological constraints on tone development. [Work supported by NIH 5F31DC8470 and NSF.]

4pSW7. Effectiveness of a perceptually weighted measure on automatic evaluation of a foreign accent in a second language. Hiroaki Kato (NICT/ATR, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan, kato@atr.jp), Shizuka Nakamura (GITI, Waseda Univ., Shinjuku-ku, Tokyo 169-0051, Japan), Shigeki Matsuda (NICT/ATR), Minoru Tsuzaki (Kyoto City Univ. Arts, Nishikyo-ku, Kyoto 610-1197, Japan), and Yoshinori Sagisaka (GITI, Waseda Univ./NICT, Tokyo, Japan)

An empirical study is carried out to achieve a computer-based methodology for evaluating a speaker's accent in a second language as an alternative to a native-speaker tutor. Its primary target is the disfluency in the temporal aspects of an English learner's speech. Conventional approaches commonly use measures based solely on the acoustic features of given speech, such as segmental duration differences between learners and reference native speakers. However, our auditory system, unlike a microphone, is not transparent: it does not send incoming acoustic signals into the brain without any treatment. Therefore, this study uses auditory perceptual characteristics as weighting factors on the conventional measure. These are the loudness of the corresponding speech segment and the magnitude of the jump in loudness between this target segment and each of the two adjacent speech segments. These factors were originally found through general psychoacoustical procedures [H. Kato et al., JASA, **101**, 2311–2322 (1997); **104**, 540–549 (1998); **111**, 387–400 (2002)], so they are applicable to any speech signal despite the difference in language. Experiments show that these weightings dramatically improve evaluation performance. The contribution of psychoacoustics to evaluation methodology of second language speech is also discussed. [Work supported by RISE project, Waseda Univ. and KAKENHI 20300069, JSPS.]

4pSW8. Effects of "catch" trials on the discrimination of American English vowels by native Japanese speakers. Takeshi Nozawa (Lang. Education Cntr., Faculty of Economics, Ritsumeikan Univ. 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan)

Native Japanese speakers' ability to discriminate American English vowels was tested in two different formats, in both of which they heard three stimuli per trial. The three stimuli in each trial were produced by three different talkers. In AXB format, subjects were to decide if the second stimulus included categorically the same vowel as the first or the third stimulus. In other words, the subjects knew that one of the three stimuli was categorically different from the other two. In the other format, the subjects heard three stimuli in each trial, and chose one stimulus that includes a categorically different vowel from the other two. Each vowel contrast was tested by different trials which contained an odd item out, and catch trials which contained three tokens each from each of the vowel contrast pairs. Subjects were visually presented with four boxes on a computer screen with the numbers 1, 2, and 3, and "No." The "No" button was provided when no differences were detected The results reveal that with catch trials included, discriminating American English vowels is a lot more difficult. The results imply the Japanese subjects' susceptibility to talker variance and inability to discriminate American English vowels categorically.

4pSW9. A twin study of speech perception and production in first and second language among Chinese children learning English as a second language. Simpson Wai-Lap Wong, Dorothy Bishop (Dept. of Experimental Psych., Univ. of Oxford, S. Parks Rd., Oxford, OX1 3UD, UK, wlswong@gmail.com), and Connie Suk-Han Ho (Univ. of Hong Kong, Hong Kong)

The faculty of language processing in our brain develops with the constraint of our genetic dispositions and also our experiences. Do the same genetic influences affect learning of phonology in L1 and L2? This study examines the genetic and environmental effects on the individual differences on L1 and L2 speech perception and production among Chinese children learning English as a second language. By employing a twin-study design, 150 pairs of monozygotic and 150 pairs of same-sex dizygotic twins aged from 4 to 11 were tested. Children's speech perception and production in both languages were assessed with an AXB speech perception task of minimal word pairs and two picture naming tasks, respectively. Children's nonverbal IQ was also measured. Scores of accuracy will be computed for each task. Analyses will be conducted to consider the relationship between accuracy in perception and production in L1 and L2, to estimate the extent of genetic contribution to speech perception, and to determine whether heritable individual differences are accounted for by a common factor, or whether different factors influence proficiency in L1 and L2.

4pSW10. Lexical frequency, orthographic information, and first-language effects on second-language pronunciation. Ron I. Thomson (Dept. of Appl. Linguist., Brock Univ., 500 Glenridge Ave., St. Catharines, ON L2S 3A1, Canada, rthomson@brocku.ca) and Talia Isaacs (McGill Univ., Montreal, QC H3A 1Y2, Canada)

In L2 speech learning, lexical frequency may play a facilitative role, whereby perception and production of sounds found in high-frequency lexical items will develop before the perception and production of the same categories found in low-frequency lexical items (see Munro and Derwing, 2008). Orthographic information may also facilitate learning by disambiguating L2 sounds (see Erdener and Burnham, 2005), particularly in known words. This study examines the role of lexical frequency, orthographic information, and a learner's L1 in the development of L2 speech perception and production. Thirty-eight Mandarin and Slavic participants were asked to repeat a word list comprising ten target English vowels, each embedded in three separate monosyllabic verbs and varying in lexical frequency. Recordings of the L2 productions were obtained under three counter-balanced conditions: (1) after hearing an auditory prompt accompanied by the written form of the word; (2) after hearing an auditory prompt with no written form provided; and (3) with no auditory prompt but the written form provided. To measure L2 performance, L1 English listeners were asked to identify the vowel they perceived in each recorded item. Results were examined to determine what influence lexical frequency and orthographic information might have had on L2 performance.

4pSW11. A closer look at perceptual epenthesis in cross-language perception. Lisa Davidson and Jason Shaw (Dept. of Linguist., New York Univ., 726 Broadway, New York, NY 10003, lisa.davidson@nyu.edu)

Previous studies have shown that listeners have difficulty distinguishing non-native CC sequences from their CVC counterparts (e.g. Japanese [ebzo]-[ebuzo], English [zgamo]-[zəgamo]). Some have argued that the phonology "repairs" non-native consonant clusters with vowel insertion ("perceptual epenthesis"), causing listeners to respond that CC and CVC sequences are the same. Production studies, however, have shown that speak-

ers produce non-native CC sequences many ways, including epenthesis, changing C1, deleting C1, and prothesis. To test whether these other repairs of non-native clusters are also difficult to distinguish, English listeners (n=37) participated in an AX discrimination task that paired C1CVCV stimuli with CəCVCV (epenthesis), əCCVCV (prothesis), CVCV (deletion), or C2CVCV (C1 change). Each of these repairs caused some difficulty and there was an interaction between repair type and manner combination (stop-stop, stop-nasal, fricative-stop, fricative-nasal). Listeners were more accurate when C1 was deleted (81% accurate), followed by epenthetic and C1 changes (both 77%), then by prothesis (59%). Furthermore, compared to previous studies testing only the epenthetic repair, presenting listeners with various repair types dramatically improves discrimination in the epenthesis condition. These results indicate that perceptual epenthesis may be a task effect and that top-down phonological influences are more complex than previously assumed. [Research supported by NSF.]

4pSW12. Native and non-native clear speech production. Rajka Smiljanic (Linguist., Univ. of Texas at Austin, 1 University Station B5100, Austin, TX 78712-0198) and Ann Bradlow (Northwestern Univ., Evanston, IL 60208)

Previous work established that both native and non-native clear speech increase intelligibility for native and proficient non-native listeners. However, non-native clear speech enhanced intelligibility less than native clear speech. In this study, we examine native and non-native conversational and clear speech productions with an eye on the differences in hyperarticulation strategies that may underlie a smaller clear speech intelligibility benefit. Results revealed a slower conversational speaking rate for non-native talkers compared with native talkers. Clear speaking rate was similar for both groups suggesting a limit in the speaking rate decrease in intelligibility enhancing clear speech. The durational distinction between tense and lax vowels was increased less in non-native speech due to the smaller speaking rate decrease in clear speech. Analyses of the stop voice onset time (VOT) showed that non-native talkers lengthened the voiced category in the negative VOT direction, while the voiceless category was lengthened less in clear speech. Finally, examination of vowel space expansion in clear speech revealed differences between the two talker groups reflecting non-native talkers' inexperience in producing some of the vowel targets. Combined, these results suggest that a complex interaction of articulatory difficulty, proficiency, and native language background determines levels of non-native clear speech intelligibility.

4pSW13. Measuring sensitivity to short-term deviations in acoustic cue correlation. Kaori Idemaru (Dept. of East Asian Lang. and Lit.s, Univ. of Oregon, Eugene, OR 97403) and Lori L. Holt (Carnegie Mellon Univ., Pittsburgh, PA 15213)

Listeners are sensitive to correlations among the multiple probabilistic acoustic cues that define speech categories. In English stop consonant productions, for example, fundamental frequency (f0) is strongly correlated with voicing. Reflecting this regularity, perception of syllables varying in voice onset time is shifted with changes in f0. Such sensitivity to the long-term regularities of a language must be balanced with enough flexibility that speech perception is able to accommodate deviations from these regularities, such as those that may arise variations from idiolect, dialect, or accented speech. The present experiments manipulate short-term acoustic cue correlations experienced in online perception to investigate the interplay between sensitivity to long-term and short-term acoustic cue correlation. Using overt categorization and eyetracking, the present results indicate that speech categorization is influenced by local shifts in acoustic cue correlations that deviate from long-term regularities of English. These experiments further examine the time course of this short-term learning and the degree to which it generalizes. The data suggest that listeners continually monitor speech input for regularity and tune online speech perception in relation to these regularities.

4pSW14. Articulatory analysis of foreign-accented speech using real-time MRI. Joseph Tepperman (Viterbi School of Eng., Univ. of Southern California, 3740 McClintock Ave., EEB 400, Los Angeles, CA 90089, tepperma@usc.edu), Erik Bresch, Yoon-Chul Kim, Louis Goldstein, Dani Byrd, Krishna Nayak, and Shrikanth Narayanan (Univ. of Southern California, Los Angeles, CA 90089)

We present the first study of nonnative English speech using real-time MRI analysis. The purpose of this study is to investigate the articulatory nature of "phonological transfer"—a speaker's systematic use of sounds from their native language (L1) when they are speaking a foreign language (L2). When a non-native speaker is prompted to produce a phoneme that does not exist in their L1, we hypothesize that their articulation of that phoneme will be colored by that of the "closest" phoneme in their L1's set, possibly to the point of substitution. With data from three native German speakers and three reference native English speakers, we compare articulation of read phoneme targets well documented as "difficult" for German speakers of English (/w/ and /dh/) with their most common substitutions (/v/ and /d/, respectively). Tracking of vocal tract organs in the MRI images reveals that the acoustic variability in a foreign accent can indeed be ascribed to the subtle articulatory influence of these close substitutions. This suggests that studies in automatic pronunciation evaluation can benefit from the use of articulatory rather than phoneme-level acoustic models. [Work supported by NIH.]

4pSW15. The perception of speech and nonspeech analogs by first- and second-language speakers. Paul Iverson, Anita Wagner, and Stuart Rosen (Res. Dept. Speech Hearing and Phonetic Sci., University College London, 2 Wakefield St., London WC1N 1PF, UK)

There are marked differences in how native speakers of Sinhala and English perceive the English /w/-/v/ distinction; Sinhala speakers who have learned English as a second language are typically near chance at identification and have less than half the discrimination sensitivity of native English speakers. This poor discrimination ability is remarkable because the acoustic cues for /w/ and /v/ are varied, being contrastive in formant frequencies, amount of frication, and amplitude envelopes. The present project explored these differences in auditory sensitivity by manipulating speech and nonspeech stimuli, with the aim of investigating how close an auditory stimulus needs to be to natural speech in order for cross-language perceptual differences to emerge. Synthesized VCV speech stimuli were created to model natural recordings. Nonspeech stimuli were created by removing dynamics from the stimuli (e.g., flat pitch and amplitude, no formant movement), producing a "buzz" that was acoustically similar to the speech stimuli. Discrimination results for these stimuli will be reported for native speakers of Sinhala and English, in order to evaluate whether cross-language specialization for speech may occur at a precategorical auditory-phonetic level of processing, or whether specialization is contingent on the stimuli being perceived as intelligible speech.

4pSW16. Proficiency instrumentation for cross language perception studies. Jared Bernstein (Pearson, 299 S. California Ave., Ste. 300, Palo Alto, CA 94306, Jared.Bernstein@Pearson.com)

Studies in cross language speech perception sometimes use self-reported values of language proficiency as an independent variable. In relation to other measurement methods, proficiency self-reports have been found susceptible to several sources of bias including demographic variables and proficiency level itself [S. Yamashita, *Six Measures of JSL Pragmatics*, U. Hawaii Press (1996)]. An inaccurate scaling of an independent variable can limit the precision of performance models that are based on that variable's relation to the dependent variable. This paper presents a convenient method (Versant tests) for measuring spoken language proficiency that locates subjects on a 60 point scale with a standard error measurement of less than 3 points, thus reliably locating subjects into 20 levels along a linear continuum of proficiency. The testing procedure takes about 15 minutes to complete. The test-retest reliability of Versant scores for spoken English and Spanish are 0.97 and 0.96, respectively, both with correlations greater than 0.85 with independent human rated tests. The development, operation, and psychometric properties of the Versant tests are presented briefly, along with examples of their use in evaluating language instruction programs and standard listening comprehension tests.

4pSW17. Why do adults vary in how categorically they rate the accuracy of children's speech? Eden Kaiser (Prog. in Linguistic, Univ. of Minnesota, 214 Nolte Ctr., 315 Pillsbury Dr. SE, Minneapolis, MN 55455, kaise113@umn.edu), Benjamin Munson (Univ. of Minnesota, Minneapolis, MN 55455), Fangfang Li (Univ. of Lethbridge, Lethbridge, AB T1K 3M4, Canada), Jeff Holliday (Ohio State Univ., Columbus, OH 43210), Mary Beckman (Ohio State Univ., Columbus, OH 43210), Jan Edwards, and Sarah Schellinger (Univ. of Wisconson-Madison, Madison, WI 53706)

In a recent experiment using continuous visual analog scales (VAS) to examine adults' perception of children's speech accuracy, listeners varied in the extent to which they categorically perceived children's English and non-English productions [Munson *et al.*, American Speech-Language-Hearing Association (2008)]. Some listeners utilized all points on the VAS line for their responses, while others grouped responses around discrete locations on the line. It is hypothesized that differences in categoricity of responses across listeners might relate to listeners attending to either categorical linguistic information (i.e., identifying phonemes in a word, which would promote more categorical labeling) or gradient indexical information (i.e., identifying the child's sex or age, which would promote more continuous labeling). If this is true, it should be possible to elicit differences in categoricity of fricative "goodness" judgments in individual listeners by priming them to listen to linguistic variables (by interleaving fricative judgment trials with trials in which they categorize the vowel spoken by the child) or indexical variables (by interleaving fricative judgment trials with trials in which they identify the child's sex). This paper reports on an experiment designed to test this. Results will help us better understand individual response patterns in cross-language speech perception experiments.

4pSW18. Perceptual studies of two phonological voicing patterns. Scott Myers (Dept. of Linguist., 1 Univ. Sta. B5100, Univ. of Texas, Austin, TX 78712, s.myers@mail.utexas.edu)

Two of the most common phonological patterns in the world's languages are final devoicing and regressive voicing assimilation. In final devoicing, there is a contrast in obstruent voicing generally, but in word-, phrase-, or syllable-final position, only voiceless obstruents are allowed (e.g., Dutch, Russian, Walloon). In regressive voicing assimilation, the first of a sequence of two obstruents is required to agree with the second in voicing (e.g., Russian, Catalan, Hebrew). These phonological patterns can be related to patterns of speech production: anticipation of the spread glottis posture in pause and laryngeal coarticulation within consonant clusters. The studies reported here investigate the perceptual consequences of these coarticulatory effects. Native speakers of English listened to words contrasting in the voicing of the final consonant (cease/seize) and identified which word they heard. The words were excised from utterance-final position or from phrase-medial position preceding a vowel, nasal, a voiced obstruent, or a voiceless obstruent. The subjects averaged about 90 percent correct identification. But they still displayed a significant bias toward identifying as voiceless the final consonant in utterance-final words as well as in words preceding a voiceless obstruent. This pattern of identification errors could be the diachronic basis of the phonological patterns.

4pSW19. The relationship between speech perception and production in second language learners. Bing Cheng (English Dept., Inst. of Intil. Studies, Xi'an Jiaotong Univ., Xi'an, China, 710049) and Yang Zhang (Univ of Minnesota, Minneapolis, MN 55455)

This study aims to examine the relationship between perception and production skills in second language learners. Thirty-nine Chinese college students who have received at least eight years of English-as-a-second-language (ESL) education in school participated in the study. The data were collected using two programs: HearSay from Communication Disorders Technology Inc., Indiana, and a perceptual test program developed by Yang Zhang at the University of Minnesota. Digital recordings of each participant's production of the individual words used in the perceptual test program were rated in the scale of 1 to 5 by two native speakers of English. Results demonstrated the existence of complex relationships between perception and production in ESL learners. Overall, the consonants demonstrated a significant positive correlation between perception and production ($p<0.01$) with voicing showing strongest effect. Vowels, on the other hand, did not show significant correlations in any of the five categories examined: tense/lax, high/low, front/back, diphthong, and vowel insertions. Four patterns were identified for the individual phonemes: (a) good perception and good production, (b) poor perception and poor production, (c) good perception and poor production, and (d) poor perception and good production. The results are discussed with respect to speech learning models and practical instructional approaches.

4pSW20. Indexical perception in Chinese: The influence of first-language (L1) Taiwanese on second-language (L2) Mandarin. Yufen Chang (2451 E. 10th St., Apt. 909, Bloomington, IN 47408, yufchang@indiana.edu)

A pilot study on indexical perception in Chinese hypothesized that L1 Taiwanese speakers' L2 Mandarin production could be detected as having a nonnative accent. If the nonnative accent was detectable, what phonetic properties (the phonetics of consonants or vowels, or the prosody) did L1 Taiwanese speakers' Mandarin exhibit such that these attributes entailed indexical information? Additionally, the study argued that L1 Mandarin speakers would show higher sensitivity to the nonnative accent than L1 Taiwanese speakers. The results have shown that only in sentence production, the sentences being either low-pass filtered or unfiltered, could the L1 Taiwanese talkers be possibly indexed correctly. It was also found that the listeners perceived more accurately when they encountered retroflex stimuli in monosyllabic word condition. Based on these findings, it was concluded that the prosody and the retroflex consonants seemed to reveal much of the talker's L1 information. However, concerning the hypothesis that L1 Mandarin speakers were more sensitive to the nonnative accent, the results did not support this. The findings in this pilot study are preliminary, given the fact that the corpus was small. A bigger-scale study is being carried out to gain more insights into indexical perception in Chinese.

4pSW21. Automatic auditory discrimination of vowels in simultaneous bilingual and monolingual speakers as measured by the mismatch negativity (MMN). Monika Molnar, Shari Baum, Linda Polka (McGill Univ., 1266 Pine Ave. W, Montreal, H3G 1A8 Canada, monika.molnar@mcgill.ca), Lucie Menard (Universite du PQ à Montréal, Montreal, Canada), and Karsten Steinhauer (McGill Univ., Montreal, H3G 1A8 Canada)

MMN responses reflect whether language users have developed long-term memory traces in response to phonemes and whether they are able to perceive small acoustic changes within speech sound categories. Subtle acoustic changes within phonemes are often irrelevant to monolingual perceivers, but can be crucial for bilingual perceivers if the acoustic change differentiates the phonemes of their two languages. In the present study, we investigated whether bilinguals are sensitive to such acoustic changes. We recorded MMN responses from monolingual (English, French) and simultaneous bilingual (English/French) adults using an auditory oddball paradigm in response to four vowels: English [u], French [u], French [y], and an acoustically-distinct (control) [y]. In line with previous findings, monolinguals were more sensitive to the phonemic status of the vowels than to the acoustic properties differentiating the sounds. Bilingual speakers revealed a different pattern; they demonstrated overall slower discrimination responses to all sounds, but showed almost equal sensitivity to phonemic and phonetic /acoustic differences. The results suggest that bilingual speakers exhibit a more flexible but less uniquely-specified perceptual pattern compared to monolingual speakers.

4pSW22. Comparison between Japanese children and adult's perception of prosodic politeness expressions. Takaaki Shochi (Div. of Cognit. Psych., Kumamoto Univ., Kurokami 2-40-1, Kumamoto City, 860-8555, Japan, shochi38@gmail.com), Donna Erickson (Showa Univ. of Music, Kawasaki-city, Kanagawa pref., 215-0021, Japan), Albert Rilliard (LIMSI-CNRS, BP 133 F-91403 Orsay cedex, France), Kaoru Sekiyama (Kumamoto Univ., Kurokami 2-40-1, Kumamoto city, 860-8555, Japan), and Veronique Auberge (Domaine Universitaire, BP46 F-38402 Saint Martin d'Heres cedex, France)

Previous work examined the contribution of audio and visual modalities for perception of Japanese social affects by adults. The results showed audio and visual information contribute to the perception of culturally encoded expressions, and show an important synergy when presented together. Multimodal presentation allows foreign adult listeners to recognize culturally encoded expressions of Japanese politeness which they cannot recognize with an audio-only stimuli. This current work analyzes the recognition performance of politeness expressions by Japanese children 13 to 14 years old.

Stimuli, based on one sentence with an affectively neutral meaning, are performed with five different expressions of politeness. Subjects listen three times to each stimulus and judge the intended message of the speaker. The stimuli are presented as audio-only, visual-only, audio-visual. Listeners rate the social status of the hearer and the degree of politeness on a nine-point scale ranging from polite to impolite. The results are analyzed to capture the relative ability of adults and children to use both modalities to recognize social affects. [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.]

4pSW23. Do you want to pronounce correctly in a foreign language? Start music lessons!. Barbara E. Pastuszek-Lipinska (School of English, Adam Mickiewicz Univ., al. Niepodleglosci 4, 61-874 Poznan, Poland)

Music education is one of the human activities that requires the integration of all human senses and the involvement of all cognitive processes: sensory, perceptual, and cognitive learning, memory, emotion, and auditory and motor processes, music has tended to figure only marginally in an approach to second-language acquisition. In the presentation will be reviewed recent research results on the influence of music lessons on second language speech processing. The presentation will draw a comparison between second language acquisition processes and first language acquisition processes. It will be also provided concise review of the possible explanations of the influence of music on second language speech processing. The review will include acoustic, phonological, neuroscientific, and psycholinguistic approaches.

4pSW24. Mandarin-English bilinguals' accented first-language (L1) vowel production. Haisheng Jiang (School of Lang. and Lit., Univ. of Aberdeen, Aberdeen, AB24 3FX, UK, h.jiang@abdn.ac.uk)

The L1 phonetic categories established in childhood may undergo modification when similar L1 and L2 sounds interact in the process of L2 learning [J. Flege, *Speech perception and linguistic experience: Issues in cross-language research* (1995)]. In this study, Mandarin vowel production by the Mandarin-English bilinguals was presented to Mandarin as well as English listeners for goodness rating. The results showed that both Mandarin-English bilinguals of high L1 use and those of low L1 use differed significantly from Mandarin monolinguals in the production of /y/, a vowel with no counterpart in English. An analysis of inter-speaker variability indicated that some individual Mandarin-English bilinguals, including both speakers of high L1 and low L1 use, were accented in the production of /y/, /aj/, and /au/. Possible acoustic properties contributing to their accentedness were identified. L2 English learning led to some Mandarin-English bilinguals carrying some English characteristics in their L1 Mandarin vowel production. This study provides further evidence for the claim that the L1 phonetic system established in childhood is susceptible to change. It contributes to the less well-studied field of L2 influence on L1.

4pSW25. Assessing phonetic talent in second language performance: The relationship between perception abilities in the native and the second language and production. Matthias Jilka (Institut of English Linguist., Univ. of Stuttgart, Heilbronner Str. 7, 70174 Stuttgart, Germany, jilka@ifla.uni-stuttgart.de), Giusy Rota (Univ. of Stuttgart, 70174 Stuttgart, Germany), and Susanne Reiterer (Univ. of Tuebingen, 72076 Tuebingen, Germany)

An extensive project that aims to assess innate phonetic talent is introduced. It investigates language performance of 103 native speakers of German in their native language, in English and (to a limited degree) in Hindi. The project's original priority is to establish a talent score that serves as the basis for a neurolinguistic study which attempts to find correlates between phonetic talent and brain anatomy/function. For this purpose tasks investigating neurobiological, psychological, and other factors influencing performance are also conducted. Innate talent is to be distinguished from superficial proficiency by limiting or eliminating the influence of interfering factors such as language experience. The test subjects are investigated with respect to different manifestations of their production (e.g., quasi-spontaneous speech, reading), perception (e.g., discrimination, identification), and imitation abilities in both German and English. A first analysis of the interactions between all perception and production scores shows a significant correlation at the level of $p < 0.05$. The correlation is especially high ($p < 0.01$) between perception tasks involving the interpretation of intonational features of both English and German and production abilities (again both in German and English), suggesting a certain cross-language predictive power.

4pSW26. Speech perception assessment and training system (SPATS-ESL) for speakers of other languages learning English. James D. Miller, Roy Sillings, Charles S. Watson, and Diane Kewley-Port (Commun. Disord. Technol., Inc., 501 N. Morton St., Sta. 215, Bloomington, IN 47404)

The SPATS software system, originally developed for the hearing-impaired, has been modified for use with ESL learners with TOFEL (pbt) scores near and well above 500. SPATS-ESL includes the identification of syllable constituents: onsets, nuclei, and codas, as well as sentence recognition. The syllable constituent tasks include the progressive introduction of increasing numbers of constituents until the learner becomes adept at the identification of 45 onsets, 28 nuclei, and 36 codas presented by eight talkers in a variety of phonetic contexts. The sentence task emphasizes increasing speed and decreasing errors in the recognition of short, meaningful sentences spoken by a variety of talkers. The sentences are presented in a background of multi-talker babble at five signal-to-noise ratios: +10, +5, 0, −5, and −10 dB. The syllable constituent and sentence tasks are interleaved throughout training. In constituent training, SPATS uses a proprietary training algorithm, Adaptive Item Selection (AIS), which automatically focuses training on individuals' items of intermediate difficulty and is independent of their language history. Proctored tests allow certification of a learner's English speech perception in relation to native-speaker performance.

4pSW27. On-line perception of lexical stress in English and Spanish: The effects of sentence intonation and vowel reduction. Marta Ortega-Llebaria (Dept. of Spanish and Portuguese, Univ. of Texas at Austin, Austin, TX 78712, ortegallebaria@att.net)

Previous research shows that speakers of stress-accent languages rely on pitch-accents to perceive word stress in sentences spoken with declarative intonation, while in unaccented sentences, like post-focal contexts, they rely on other cues, i.e., duration in Spanish or vowel reduction in English. However, there is no experimental evidence on the effect that sentence intonation has in the "on-line" perception of word stress across languages. This experiment examines whether listeners detect word stress faster in unaccented stretches of speech that are preceded by a landmark in the sentence melody, i.e., the focal pitch-accent that always precedes a post-focal sentence, than in unaccented contexts not preceded by a landmark, i.e., reporting clauses. Results show that Spanish listeners identified target words at the beginning and end of post-focal contexts with similar reaction times, while in reporting clauses, target words placed at the end were identified faster than those placed at the beginning. Preliminary results for English show no significant differences. Thus, Spanish listeners re-weight cues to word stress on-line taking into account the patterns of sentence intonation. Sentence melody does not have such a strong effect in English, possibly because vowel reduction is a sufficient cue to an effective perception of word stress.

4pSW28. A cross-language study of compensatory response to formant-shifted feedback. Takashi Mitsuya, Ewen N. MacDonald (Dept. of Psychol., Queen's Univ., Kingston, Canada), David W. Purcell (Univ. Western Ontario, London, Canada), and Kevin G. Munhall (Queen's Univ., Kingston, Canada)

Previous experiments in speech motor learning have demonstrated that the perception of our own voice while we speak plays a role in the control of fundamental and formant frequencies and vocal amplitude. When feedback is changed in real time, subjects alter speech production in attempt to compensate for the perturbation. By testing Japanese talkers in their native and a less familiar language (as well as English-speaking controls), we examine how this perception-production process is influenced by language. In the first study, native Japanese speakers produced an English word with formant-shifted feedback. In the second experiment, native Japanese speakers produced a Japanese syllable with altered feedback and produced an

English word that contained a similar vowel with normal feedback. The results were compared with data from English controls and suggest that the compensatory behavior is not language dependent.

4pSW29. Linguistic experience's influence on foreign accent detection in short, slightly accented speech. Hanyong Park (Speech Res. Lab., Indiana Univ., 1101 E. 10th St., Bloomington, IN 47405, hanypark@indiana.edu)

This study examined how a listener's linguistic experience affects the detection of foreign accent in short, slightly accented speech. Two factors were considered: Age of arrival in a second-language (L2) speaking country (AOA), and length of residence in that country (LOR). Two listener groups with the same native language (L1: Korean) but with different AOA and LOR judged the nativeness of short stimuli produced by 4 Korean-English bilinguals and 2 native speakers of American English. The stimulus length ranged from the vowel /a/ to monosyllabic and disyllabic English words. To investigate the listeners' sensitivity patterns to different linguistic structure, the monosyllabic corpus included stimuli having both natural (i.e., CV) and unnatural syllable structures (i.e., CCV, CVC, and CCVC) as well as various segments in terms of Korean phonotactics. Results show that: (1) more experienced with L2 led to higher sensitivity to a foreign accent; (2) AOA affects foreign accent detection more than LOR; and (3) the non-native listeners were not sensitive to different linguistic structures, except in different CCV types by the more experienced listeners. The results suggest that a listener's sensitivity to a foreign accent develops up to a certain degree, according to L2 experience.

4pSW30. Asymmetries in the mismatch negativity response to vowels by French, English, and bilingual adults: Evidence for a language-universal bias. Linda Polka, Monika Molnar, Shari Baum (School of Commun. Sci. and Disord., McGill Univ., 1266 Pine Ave. W., Montreal, QC Canada, linda.polka@McGill.ca), Lucie Menard (Universite du Qubec à Montréal, QC Canada), and Karsten Steinhauer (McGill Univ., Montreal, QC, Canada)

In infants, discrimination of a vowel change presented in one direction is often significantly better compared to the same change presented in the reverse direction. These directional asymmetries reveal a language-universal bias favoring vowels with extreme articulatory-acoustic properties (peripheral in F1/F2 vowel space). In adults, asymmetries are observed for non-native but not for native vowel contrasts. To examine neurophysiological correlates of these asymmetries we recorded MMN responses from monolingual (English, French) and simultaneous bilingual (English/French) adults using an oddball paradigm with four vowels: English [u], French [u], French [y], and an acoustically-distinct [y]. All vowels were tested in four conditions with each vowel serving as deviant and standard. Within each vowel pair, MMN responses were larger and earlier when the deviant vowel was more peripheral. This pattern was consistent across the language groups and was observed for within-category (within [u]; within [y]) and for cross-category vowel pairs ([u] versus [y]). Findings indicate that a bias favoring peripheral vowels is retained in the neural pre-attentive processing of vowels in adults. As in infant behavioral findings, this bias is independent of the functional status of the vowel pair. Implications for the natural referent vowel model [Polka and Bohn (2003)] will be discussed.

4pSW31. The effect of accent on toddlers' story comprehension and word recognition. Brittan A. Barker and Lindsay E. Meyer (Dept. of COMD, Louisiana State Univ., 63 Hatcher Hall, Baton Rouge, LA 70803)

The current study sought to enrich the developmental research and contribute to the ever-growing knowledge about talker-specific spoken language processing by examining the affect of accent on toddlers' ability to comprehend a story. A total of 24 children aged 30–42 months participated in the first study employing a between-subjects design. Talker accent (native versus non-native) served as the independent variable and story comprehension accuracy served as the dependent variable. It was predicted that the children listening to the story in their native accent would perform significantly better on the comprehension task than those listening to the non-native accent. The hypothesis was not supported; there was no significant difference between the listeners' performance [$t(22)=.169$, $p>.05$]. An ongoing, follow-up study was conducted to further explore the surprising results. The affect of accent on word recognition skills in these same children is currently being tested. The complete data set of the second study will be presented. It is hypothesized that the toddlers will have significantly more difficulty recognizing the words spoken in a non-native accent. These results would echo previous talker-specific word recognition results [Ryalls and Pisoni (1997)] and suggest that talker-specific information may affect low-level speech perception but not higher-level language comprehension.

4pSW32. Cross-linguistic variation in language similarity classification. MaryAnn Walter (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, m-walter@northwestern.edu)

This study aims at identifying factors that make language sound structures seem more or less similar to English, and how those similarity judgments change according to the listener's native language. Listeners from four different native language groups (English, Mandarin Chinese, Korean, and Turkish) sorted a group of 17 genetically and geographically diverse languages in terms of their sound-based distance from English. Placements of individual languages were analyzed, as well as groupings of similarly ranked languages and correlations among overall ranking structures of the different groups. Bilingual listeners exhibit more variability in their rankings than monolingual English speakers, rank their own language as less similar to English than other groups do, and rank languages of neighboring groups the least similar to it of all. Ranking correlations between language groups are significant, varying somewhat in magnitude depending on geographical proximity and typological/genetic relatedness of the listener group languages. This reflects the presence of consistent groupings within ranking structures for all language groups, which depend on sound-based factors such as the presence of perceptually salient speech sounds. These results enable predictions about relative intelligibility among international English users, native and non-native. [Work supported by a Mellon Postdoctoral Fellowship awarded by the Woodrow Wilson Foundation.]

4pSW33. Acoustic similarities between front rounded and back unrounded vowels as evidenced by French /ø/ and /u/ produced by Japanese-speaking learners. Takeki Kamiyama (Laboratoire de Phonetique et Phonologie (UMR 7018) CNRS/Sorbonne Nouvelle, 19, rue des Bernardins, 75005 Paris, France, Takeki.Kamiyama@univ-paris3.fr)

French high back rounded /u/ is characterized by a concentration of energy in the low frequency zone ($<$ 1000 Hz) due to the grouping of the first two formants, and midhigh front rounded /ø/ by a balanced distribution of formants, with F2 located around 1500 Hz. Japanese-speaking learners of French (JSL), who have difficulty differentiating /ø/ and /u/ both in perception and production, tend to produce, for both, Japanese-like /u/, which is fronted and less rounded, with a formant structure similar to that of French /ø/. Our perception experiment using 18 tokens each of /u y ø/ produced by five JSL shows that the 16 native speakers of French (NF) tested perceived mainly /ø/ when they heard those stimuli intended as /u/ by JSL but produced with a high F2 between 1100 and 1600 Hz, with a mean goodness rating of 2–4.5 out of 5 for /ø/. Another perception test conducted with stimuli synthesized using Maeda's articulatory synthesis (VTCalcs) shows that the 16 NF examined tended to identify vowels synthesized as front rounded vowels and back unrounded vowels both as front unrounded ones (/ø/ and /æ/, respectively), which indicates the acoustic similarities of both types of vowels.

4pSW34. Relation of perception training to production of codas in English as a second language. Teresa Lopez-Soto (Dept. of English Lang., Univ. of Seville, Palos de la Frontera, Sevilla 41004, Spain and Indiana Univ., Bloomington, IN 47405) and Diane Kewley-Port (Indiana Univ., Bloomington, IN 47405)

A preliminary study has been conducted to discover whether moderate amounts of speech perception training improve accurate production even though production is not trained. The study recruited one group of eight Spanish adults who had resided less than 10 years in the USA. A set of 13 word-final English consonants was selected for training from a SPATS software module. On days 1 and 5, the group participated in both perception and production tasks with the 13 codas (pre- and post-tests). On days 2–4, the group trained with feedback for 1 h mostly with VC syllables and occasionally with sentences (speech recorded from multiple talkers). Results show: (1) with 3 h training, Spanish listeners' perception improved significantly across the 13 codas, with greater improvement on consonant clusters than on singletons; (2) for consonants not accurately produced in the pretest, many substantially improved after only perception training; (3) several consonants

with large gains in perception also showed the large improvements in production. The results from this study suggest that training only with perception can improve speech production. Experience with this protocol lays the groundwork for a series of studies to examine how perception and production are linked in learning a new language.

4pSW35. Confusion direction differences in second language production and perception. Yen-Chen Hao and Kenneth de Jong (Dept. of Linguist., Indiana Univ., 322 Memorial Hall, 1021 E. 3rd St., Bloomington, IN 47405)

This study examines differences between confusions found in productions and perceptions of learners of English. Twenty Korean EFL learners engaged in three tasks involving obstruents placed in different prosodic positions: (a) identification of native English productions, (b) reading from orthographic prompts, and (c) mimicry of native English productions. Recordings of reading and mimicry were presented to 50 native English listeners for identification. This paper compares patterns of errors found for 10 intervocalic obstruents before and after a stress, since previous studies showed that Korean does not exhibit stress-induced differences in consonant allophones. Similarity estimates using Luce's similarity choice model were regressed across the two intervocalic positions. We found robust correlations despite allophonic differences in English, suggesting a component of L1 transfer in all three tasks. Examining bias parameters, however, revealed systematic differences in the direction of the resulting errors, which is task-dependent. Identification and mimicry tended to underestimate allophonic shifts due to stress, and so to create more voiceless to voiced errors in post-stress environments. Reading productions exhibited error directions in exactly the opposite directions, suggesting Korean learners produced the stress but not the corresponding allophonic variations. These patterns indicate very different error outcomes in production and perception.

4pSW36. Maintenance of /e-ɛ/ in word-final position as a phonemic and morphemic contrast in Canadian French. Franzo Law, II and Winifred Strange (Program in Speech-Lang.-Hearing Sci., CUNY Graduate Ctr., 365 5th Ave., New York, NY 10016, flaw@gc.cuny.edu)

Many dialects of French have merged /e-ɛ/ to /e/ in word-final context. The present study investigated the stability of this contrast in Canadian French. Productions of four Canadian French-dominant speakers (two monolingual, two bilingual) were recorded and analyzed. Real (e.g., "thé") and nonsense (e.g., "gispais") words ending with /e/ or /ɛ/ were used, as well as real and nonsense verbs distinguished morphosyntactically by the same vowel contrast (e.g., first person singular future "parlerai" versus first person singular conditional "parlerais"), all embedded in carrier sentences. Results showed that participants maintained spectral and duration distinction for the vowel contrast when preceded by labial, coronal, and back stops in real and nonsense words. All participants showed varying degrees of coarticulation for /e/ when preceded by /r/; /e/ was spectrally lower than in other contexts. Three of the four participants maintained a stable distinction in morphosyntactic context: monolinguals exhibited the best retention of the distinction, whereas the bilinguals had partially or completely overlapping distributions, merging /e-ɛ/ to /ɛ/, not /e/. These patterns suggest difficulties in perception of this contrast for English late L2 learners of French, due to the fact that this contrast is not phonotactically possible in English. [Work supported by NIH F31DC008075.]

4pSW37. On prosodic variation and pronunciation aptitude: A case study of German speakers of English. Volha Anufryk and Grzegorz Dogil (Azenbergstr. 12, 70174 Stuttgart, Germany)

The present study investigates prosodic variation, as realized by L1 German speakers of varying pronunciation aptitude (below-average, average, and above-average) in comparison with native speakers of English. The results demonstrate the frequency of distribution of various pitch contours in read and spontaneous speech samples, English and German, on both the phonological (ToBI pitch accents and boundary tones) and phonetic levels. The latter is analyzed using a parametric intonation model (Moehler 2001). On the phonological level the rising contours, i.e., boundary tones and pitch accents, have a wider distribution in German as well as English productions of the German subjects as compared to the L1 English speakers. In German the percentages in all the aptitude groups are almost equal covering about 50% of all contours. Their English language realizations, on the other hand, are marked by a scarcer distribution of the rising pattern, with the highest percentages in the below-average and average groups and lowest by the above-average speakers. Consequently, there is a clearcut tendency for the native pattern in the German speakers' samples. Nonetheless there is evidence for accommodation to the English variation pattern, which is strongest in the above-average aptitude group.

4pSW38. Native and non-native vocal imitation by vocalists and nonvocalists. Terry L. Gottfried and LaDonna Hayden (Dept. of Psychol., Lawrence Univ., P.O. Box 599, Appleton, WI 54912, gottfrit@lawrence.edu)

Several studies [e.g., Slevc and Miyaki, Psychol. Sci. **17**, 675–681 (2006)] have demonstrated that musical ability predicts phonological accuracy in second language learning. The present study examined whether vocal musicians' diction training affected their immediate imitation of familiar (English nonwords) and unfamiliar (Mandarin) syllables. Participants were 19 nonmusic majors (recruited from introductory linguistics) and 19 vocal music majors (who had completed three of four vocal diction classes). All participants heard a list of eight English nonwords and eight Mandarin words from a recording and repeated each word or nonword to the best of their ability. These recordings were rated (by two bilingual judges) on the overall accuracy of word repetition and vowel and consonant accuracy. Participants, regardless of group, repeated English nonwords more accurately than Mandarin words. Vocalists repeated both English nonwords and Mandarin words more accurately than the nonvocalists. This could indicate that vocalists' experience in diction training improved their ability to repeat any speech sound. Acoustical analyses of these vocal imitations are currently being conducted, and future studies will examine whether musicians without diction training are better than nonmusicians in vocal imitation.

4pSW39. Gestural drift in Spanish-English speakers. Stephen Tobin (Dept. of Psych., Univ. of Connecticut, 406 Babbidge Rd., Unit 1020, Storrs, CT 06269-1020, stephen.tobin@uconn.edu)

Voice onset times (VOTs) and second vowel formants (F2s) of native speakers of Spanish were examined to extend findings on perceptually guided changes in speech production [M. L. Sancier and C. A. Fowler, J. Phon. **25**, 421–426 (1995)]. Participants were recorded producing English and Spanish sentences after stays in the U.S. and in Spanish-speaking countries. Participants' English and Spanish VOTs were both significantly shorter after a stay in a Spanish-speaking country than after a stay in the U.S., supporting previous findings. While S1's F2s were significantly lower following exposure to Spanish, as expected based on earlier results [A.R. Bradlow, J. Acoust. Soc. Am. **97**(3), 1916–1924 (1995)], S2's F2s increased. We attribute this effect to gender, pending further data acquisition. We propose that not only inter-gestural timing but also the gestural targets themselves are affected by ambient language. The systematic effect of language exposure on F2 is consistent with language-specific articulatory settings [B. Honikman (1964); Gick et al., Phonetica **61**, 220–233 (2004)].

4pSW40. Vowel duration in Mexican heritage English. Kenneth Konopka and Janet Pierrehumbert (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, k-konopka@northwestern.edu)

Mexican heritage English (MHE) speakers in Chicago comprise a unique population of native English speakers; their speech is often regarded by non-ethnic (NE) speakers as Spanish-accented English in spite of the fact that they may have low or no proficiency in Spanish. Vowels were analyzed from recordings of wordlist readings by MHE, NE, and adult Mexican ESL speakers (L2E) from a single Chicago community ($N=15$F, 10F, and 12F, respectively). The analyses indicate that MHE vowel durations correspond to NE durations for the dimoraic (long) vowels, but correspond to the Spanish-accented vowels of L2E for the monomoraic (short) vowels. Thus the range of duration between short and long vowels is compressed for MHE vowels, reflecting the NE norm for long vowels, and the Spanish-influenced minimum for short vowels. Sociophoneticians often rely on vowel formant frequency plots to provide a quantitative measure of dialect identity that may correlate with a variety of regional and social factors. The current study provides evidence that vowel duration is a phonological contrast in the MHE dialect that reflects the influence of language contact on vowel properties that are not captured by vowel plots alone.

4pSW41. The perception of intermediate phonological relationships. Kathleen Currie Hall (Dept. of Linguist., Ohio State U., 1712 Neil Ave, Columbus, OH 43210, kchall@ling.osu.edu)

It has been shown that pairs of segments that are allophonic in a language are perceived as being more similar than pairs that are contrastive in a language [Boomershine et al., (2008)]. There is also evidence that neutralized contrasts in a language are perceived as more similar than non-neutralized contrasts [Hume and Johnson (2003)]. Third, there is evidence that phonological relationships should be defined along a continuum of predictability, rather than as a categorical distinction between "allophonic" and "contrastive" [Hall (2008)]. In combination, these facts predict that pairs of segments that fall along a cline of predictability of distribution should also fall along a cline of perceived similarity. This paper presents results of a perception experiment that tests this prediction by examining the perceived similarity of four pairs of sounds in German: (1) [t]-[tʃ], which is almost fully contrastive (unpredictably distributed); (2) [t]-[d] and (3) [s]-[ʃ], which are each partially contrastive (partially predictably distributed); and (4) [x]-[ç], which is almost fully allophonic (completely predictably distributed). If the notion of a cline of predictability is correct, these four pairs will align themselves along a cline of similarity with [t]-[tʃ] being rated as the most distinct and [x]-[ç] as the most similar.

FRIDAY MORNING, 22 MAY 2009 SKY BRIDGE AUDITORIUM, 8:45 TO 10:10 A.M.

Session 5aSWa

Speech Workshop: New Perspectives on Models of Cross-Language Speech Perception

Ann R. Bradlow, Chair
Dept. of Linguistics, Northwestern Univ., Evanston, IL 60208

Chair's Introduction—8:45

Invited Papers

9:00

5aSWa1. Accounting for the accented perception of vowels: Universal preferences and language-specific biases. Ocke-Schwen Bohn (English Dept., Arhus Univ., J.-C.-Skous Vej 5, DK-8000 Arhus C, Denmark) and Linda Polka (School of Commun. Sci. and Disord., Montreal, PQ, H3G1A8, Canada)

Strange things happen in cross-language and second-language vowel perception: Nave non-native listeners have been reported to rely on acoustic properties which are nonfunctional in their L1 and dysfunctional for the perception of non-native vowels; naïve non-native listeners' perception is guided by a preference for vowels that are peripheral in the articulatory/acoustic vowel space; and, in general, naïve non-native listeners' perception is not well predicted by comparative analyses of vowels of the native and the non-native language. This presentation reviews the accented perception of vowels by focusing on two forces which shape non-native vowel perception: universal perceptual preferences which non-native listeners (and infants) bring to the task of vowel perception, and perceptual biases which non-native listeners transfer from their native to the non-native language. Strange and her colleagues have shown that these biases cannot be predicted from acoustic comparisons; rather, they have to be examined directly through assessments of the perceived cross-language similarity of vowels. This presentation addresses several of the still unresolved questions regarding the design and the interpretation of perceptual assimilation tasks used to account for the accented perception of vowels. [Work supported by Danish Research Council for the Humanities, Canadian Natural Sciences and Engineering Research Council.]

9:35

5aSWa2. Articulating the Perceptual Assimilation Model (PAM): Perceptual assimilation in relation to articulatory organs and their constriction gestures. Catherine T. Best (MARCS Auditory Labs, Univ. Western Sydney, Locked Bag 1797, Penrith NSW 1797, Australia, and Haskins Labs, 300 George St., New Haven, CT 06511, c.best@uws.edu.au), Louis Goldstein (Univ. Southern Calif., Los Angeles, CA 90089-1693), Michael D. Tyler (Univ. Western Sydney, Penrith NSW 1797, Australia), and Hosung Nam (Haskins Labs, New Haven, CT 06511)

A core premise of the Perceptual Assimilation Model of non-native speech perception (PAM) [Best (1995); Best & Tyler (2007)] is that adults perceive unfamiliar non-native phones in terms of articulatory similarities/dissimilarities to native phonemes and contrasts. The implied attunement to native speech emerges early: As infants begin to discern the articulatory organization of native speech, language-specific effects in non-native speech perception appear (≈6–10 months). Given that non-native phones, by definition, deviate phonetically from native ones, how can we characterize articulatory similarity in concrete, testable ways? The Articulatory Organ Hypothesis (AO) [Studdert-Kennedy & Goldstein (2003); Goldstein & Fowler (2003)] offers a possible approach, positing that infants decompose the oral-facial system into distinct articulatory organs (e.g., lips, tongue tip, tongue dorsum) and are sensitive to their actions in producing vocal tract constrictions. Thus, between-organ contrasts should be easily perceived/learned by infants and adults, whereas detection of within-organ contrasts must become attuned to the distribution of differing constriction locations/types by that organ in input speech. We discuss articulatory, attunement modeling, and perceptual evidence consistent with these predictions, and present a revised version of PAM that incorporates the AO Hypothesis and related principles of articulatory phonology [Browman & Goldstein (1991)]. [Work supported by NIH.]

Session 5aSWb

Speech Workshop: Research and Applications to Second Language (L2) Speech Perception

Susan G. Guion, Chair
Dept. of Linguistics, Univ. of Oregon, Eugene, OR 97403-1290

Invited Papers

10:30

5aSWb1. Accent and intelligibility from an applied perspective. Murray J. Munro (Dept. of Linguist., Simon Fraser Univ., 8888 University Dr. Burnaby, BC V5A 1S6, Canada)

Listeners are remarkably sensitive to non-native patterns of speech, whether they are presented with full sentence productions, or with very short or severely degraded speech, including noisy, filtered, and temporally disrupted utterances. Furthermore, even phonetically unsophisticated listeners can reliably scale accents. From the standpoint of second language (L2) users, speaking with a detectible accent has important social consequences. One is that L2 speech is sometimes less intelligible or may require listeners to allocate more processing resources than does native speech. However, evidence also indicates that some salient phonological markers of L2 status have little or no impact on listener comprehension. Distinguishing these from accent features that reduce intelligibility is a critical concern in language pedagogy. Another consequence noted in empirical research is that accent stereotyping and negative social evaluation of accented speakers are linked to discrimination in remuneration, employment, and services. They may also be implicated in the exploitation of L2 speakers through questionable claims about the value of "accent reduction." This synthesis of previous and new research findings highlights a number of issues concerning methodology and interpretation in L2 speech research that are relevant to language teaching and assessment, and to human rights litigation. [Research supported by SSHRC.]

11:05

5aSWb2. The role of linguistic experience in lexical recognition. Andrea Weber (Max Planck Inst. for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands, andrea.weber@mpi.nl)

Lexical recognition is typically slower in L2 than in L1. Part of the difficulty comes from a not precise enough processing of L2 phonemes. Consequently, L2 listeners fail to eliminate candidate words that L1 listeners can exclude from competing for recognition. For instance, the inability to distinguish /r/ from /l/ in *rocket* and *locker* makes for Japanese listeners both words possible candidates when hearing their onset (e.g., Cutler, Weber, and Otake, 2006). The L2 disadvantage can, however, be dispelled: For L2 listeners, but not L1 listeners, L2 speech from a non-native talker with the same language background is known to be as intelligible as L2 speech from a native talker (e.g., Bent and Bradlow, 2003). A reason for this may be that L2 listeners have ample experience with segmental deviations that are characteristic for their own accent. On this account, only phonemic deviations that are typical for the listeners' own accent will cause spurious lexical activation in L2 listening (e.g., English *magic* pronounced as *megic* for Dutch listeners). In this talk, I will present evidence from cross-modal priming studies with a variety of L2 listener groups, showing how the processing of phonemic deviations is accent-specific but withstands fine phonetic differences.

11:40

5aSWb3. Designing the foreign language learning environment: From basic research towards product development. Reiko Akahane-Yamada (ATR Learning Technol. Corp., 2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan, yamada@atr-lt.jp)

In order to examine the acquisition of English by native speakers of Japanese, a series of training studies were conducted under various environments, that is, in the laboratory, at schools, or in virtual spaces over the Internet, and learners of various ages participated. The results demonstrated that web-based training methods using computers can improve even adult learners' ability to perceive and produce L2, and that the acquisition of phonological categories plays an important role in the language learning. Based on these results, an L2 training system, dubbed ATR CALL BRIX, was developed by putting emphasis on speech learning. The system is a collection of training tools for speech perception, production, and comprehension. Each component focuses on the acoustic-phonetic, prosodic, lexical, or semantic decoding level of spoken language. Speech analysis and pronunciation evaluation tools are also provided. The target users vary from children to adults and from beginners to advanced learners. The content is designed with Learning Object concept and each component is adjustable on the learner characteristics and curricula. The system is already in the market. It is suggested that the cross-language and L2 acquisition studies are readily applicable to designing actual foreign language learning environment. [Work supported by JSPS.]

Session 5pSWa

Speech Workshop: Cross-Language Speech Perception and Linguistic Experience: Poster Session B

Rajka Smiljanic, Chair
Dept. of Linguistics, University of Texas at Austin, Austin, TX 78712-0198

Contributed Papers

All posters will be on display from 1:30 p.m. to 3:00 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 2:15 p.m. and contributors of even-numbered papers will be at their posters from 2:15 p.m. to 3:00 p.m.

5pSWa1. Foreign accented speech: Energetic or informational masking? Lauren Calandruccio, Christina Yuen, Sumitrajit Dhar (The Roxelyn and Richard Pepper Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 N. Campus Dr., Evanston, IL 60208, lauren.calandruccio@gmail.com), and Ann Bradlow (Northwestern Univ., Evanston, IL 60208)

Normal-hearing monolingual listeners whose native language is English experience a release from masking when the competing speech stimuli are spoken in a language other than English [e.g., K. J. Van Engen and A. R. Bradlow, J. Acoust. Soc. Am. **121**, 519 (2007)]. It is unclear whether this release is due to the fact that the masking speech is not spectrally matched to the target speech (energetic influence), or that listeners are unable to understand the masking speech and therefore find it less "distracting" when spoken in a non-native language (informational influence). This study investigates listeners' recognition of English speech presented in the presence of a continuum of five two-talker babble maskers. The two-talker maskers (all created using male voices) include English babble, Mandarin-accented English babble with high intelligibility, Mandarin-accented English babble with moderate intelligibility, Mandarin-accented English babble with low intelligibility, and Mandarin babble. We hypothesize that, due to increased informational masking across the accented English continuum (from unaccented to heavily accented) in the masker speech, we will observe a increase in the target English speech intelligibility. These data will provide insight into the balance of energetic and informational influences on the release in masking observed for monolingual listeners when listening to English in a noncompeting masker language.

5pSWa2. Perceptual training really matters: evidence of a study with English as a foreign language (EFL) vowels. Denize Nobre-Oliveira (Federal Ctr. of Technolog. Education of Santa Catarina, Av. Mauro Ramos 950, Florianopolis, SC, Brazil, 88020-300, denizenobre@yahoo.com.br) and Andreia S. Rauber (Univ. of Minho, Campus de Gualtar, Braga, Portugal 4715-057)

This study aimed at: (i) investigating the perception of English vowels by Brazilian English as a foreign language (EFL) highly proficient speakers, and, based on the findings; (ii) designing and testing perceptual training tasks that could possibly minimize perception difficulties by EFL learners. Two experiments were conducted. Experiment 1 tested vowel perception by means of an identification test elaborated with synthetic stimuli. The results showed that the vowels most often misidentified were /i/, /I/, /E/, /æ/, /U/, and /u/, and that duration does not seem to play a role on the perception of each vowel, suggesting that participants rely primarily on spectral cues. Experiment 2 then tested the effect of perceptual training of the difficult vowels by means of exposure to both natural and synthetic stimuli (with spectral manipulation). The training on these vowels was given over a 3-week period to 29 Brazilian EFL learners, which were divided into the natural stimuli (NatS) group and the synthesized stimuli (SynS) group. The results show that: (i) both experimental groups improved significantly after training; (ii) there was more improvement in the SynS group than in the NatS group; and (iii) knowledge acquired with artificially enhanced stimuli was transferred to stimuli produced naturally. [Work supported by CAPES (Committee For Postgraduate Courses in Higher Education, Brazilian Ministry of Education) grants to both authors.]

5pSWa3. Perception of Arabic and Japanese vowel length contrasts by native versus non-native Japanese listeners. Kimiko Tsukada (Dept. of Int. Studies, Macquarie Univ., NSW 2109, Australia ktsukada@mq.edu.au) and Yukari Hirata (Colgate Univ., Hamilton, NY 13346)

The perception of short versus long vowel contrasts in Arabic and Japanese was examined for native Japanese [(NJ), $n=5$] and non-native Japanese [(NNJ), $n=5$] listeners. The length contrast is phonemic in both languages. The question addressed was whether experience with Japanese helps NNJs to process the length contrast in Arabic. The listeners' discrimination accuracy was assessed in an AXB test. The stimuli were CV_1C words in Arabic and CV_1CV words in Japanese (where V_1 was either short or long). As expected, NJs were more accurate than NNJs in discriminating the Japanese contrast. While NJs made more errors in Arabic than Japanese, NNJs showed a similar level of discrimination accuracy for the two languages. Finally, the number of errors for the Arabic stimuli was comparable for NJs and NNJs. Thus, the between-group difference existed for the Japanese but not for the Arabic contrast. The possibility that familiarity with phonetic characteristics of a recently acquired language may influence the processing of sounds in an unknown language with similar characteristics to the target language will be discussed. Native Arabic perception is currently under investigation. [Work supported by Macquarie University New Staff grant.]

5pSWa4. Perception of English voiceless alveolar and postalveolar fricatives by Korean speakers. Yunju Suh (Dept. of Linguist., Stony Brook Univ., Stony Brook, NY 11794, yunjusuh@gmail.com)

Korean lacks place contrast of coronal fricatives before i. Yet three alveopalatal fricatives, fortis, lenis, and labialized, occur in this context. Fortis and lenis alveopalatals are intermediate between English alveolar and postalveolar in spectral peak location, though fortis has peak at higher frequencies, and thus is closer to English s. Peak location of labialized alveopalatal is the lowest, largely overlapping with that of English ʃ. The presence of L1 contrast among alveopalatals seems to facilitate Korean listeners' perception of English si-ʃi contrast, especially that between plain and labialized alveopalatals. Korean listeners were presented with eight-step continuum of English si (step 1) to ʃi (step 8), and asked to identify them with an L1 category. The percentage of fortis answer gradually decreased from steps 1 to 4, as lenis answer gradually increased. Labialized answer appeared below 3% until step 4, and increased above 68% at step 5, indicating categorical change of the percept from plain to labialized alveopalatals. There was no categorical change from fortis to lenis, presumably because Korean fortis-lenis contrast involves complex acoustic cues such as vowel f0, frication duration, and amplitude difference between harmonics, in addition to frication noise frequency.

5pSWa5. Electrophysiological indices of vowel discrimination in late bilinguals. Carol A. Tessel, Arild Hestvik, Dolors Girbau, Richard G. Schwartz, and Valerie L. Shafer (The Graduate Ctr.-CUNY, 365 5th Ave., Speech and Hearing-7th floor, New York, NY 10016, ctessel@gc.cuny.edu)

The purposes of the current study are: (1) To investigate discrimination of a vowel contrast not found in Spanish by late learners of English with Spanish as a first language; (2) to assess whether the use of natural consonant-vowel-consonant stimuli and multiple exemplars will show the same pattern of results as found in studies using synthetic stimuli; and (3) examine whether better speech perception as examined by the event-related potential, mismatch negativity (MMN) correlates with greater language usage as measured by a language background questionnaire. The results indicate that late-bilinguals are slower at discriminating the vowel contrast than the English monolinguals, as indexed by MMN. Monolingual English listeners showed significant MMN from 200 to 300 ms, whereas for late learners of English, the MMN was significant 50 ms later between 250 and 300 ms. Both groups showed excellent behavioral discrimination of the vowel contrast. Results also suggest that vowel category reshaping is less flexible in adult learners of a second language. The results will be discussed in relation to language usage.

5pSWa6. Allophonic alternations influence non-native perception of stress. Christine E. Shea and Suzanne Curtin (Dept. of Linguist., Univ. of Calgary, 2500 University Dr. NW, Calgary, AB T2K ON4, Canada)

We examined the identification of stressed syllables by adult L2 Spanish learners to see if it is influenced by an allophonic alternation driven by word position and stress. We utilized the Spanish voiced stop-approximant alternation, where stops occur in word onsets and stressed-syllable onsets. If L2 learners track the distribution of this alternation, they should link stops to stressed syllables in word onset position and approximants to unstressed, word medial position. Low- and Intermediate-level L1 English/L2 Spanish learners, Native Spanish and monolingual English speakers listened to a series of CVCV nonce words and determined which syllable they perceived as stressed. In Experiment 1, we crossed onset allophone and vowel stress. In Experiment 2, we alternated the onset allophone and held the vowel steady. Our results show that less experienced groups were more likely to perceive stressed vowels and approximant onset syllables as stressed. This suggests that learning the interplay between allophonic distributions and their conditioning factors is possible with experience. L2 learners track distributions in the input and this, in turn, influences their perception of other properties in the language, in this case, syllable stress. Native language distributions and target language proficiency play a role in this process.

5pSWa7. Lexical encoding of the second language vowel length contrasts. Jeong-Im Han (Dept. of English, Univ. of Konkuk, Hwayang-dong 1, Gwangjin-gu, Seoul, 143-701, Korea, jhan@konkuk.ac.kr)

Given the previous results for the L2 learners' processing difficulties of suprasegmentals as compared to segmentals, this study tests whether L2 late learners lack a proper phonological representation of suprasegmentals they could use to encode those contrasts. In Experiment 1, two groups of Korean late learners of Japanese (beginner versus advanced) as well as Japanese controls took a simple AX discrimination of the nonword pairs that varied in vowel length only, and showed that both groups of Korean learners had no difficulty to perceive the acoustic differences of the long versus short vowels in Japanese. However, in Experiment 2, when the same groups of learners participated in a more demanding lexical decision task using word-nonword pairs with vowel length differences, they had much difficulty in the use of such vowel length to access the lexicon. Even though there was significant improvement between the beginner and the advanced groups, there were still significant differences between the advanced group and the Japanese natives. These results suggest that L2 late learners might have a true processing deficits for suprasegmental properties in their phonological representation which were not in their L1, which cannot be easily eliminated with a significant exposure to L2.

5pSWa8. Perception of complex word-initial onsets in English by native speakers of Japanese. Mieko Sperbeck (Dept. of Linguist., City Univ. of New York, the Graduate Ctr., 365 Fifth Ave., New York, NY 10016, msperbeck@gc.cuny.edu)

Past studies have shown that English L2 learners whose native languages have relatively simple syllable structure have different degrees of difficulty in producing complex syllable and word onsets in English. For instance, more marked clusters (e.g., /bj/) were harder to produce than the less marked ones (e.g., /pj/) for Japanese learners of English [E. Broselow and D. Finer, Second Language Research 7, 35–59 (1991)]. However, what is unknown is whether such difficulties reflect problems in perceiving complex syllable structures. The current study tested Japanese L2 learners and American English controls in a categorial ABX discrimination test of eight contrasts between nonsense words with consonant cluster onsets CC(C)VCV versus CəC(C)VCV sequences (e.g., /spani/versus/səpani/) and included /sp, sk, pl, bl, kl, gl, spl, skl/ clusters. Words were imbedded in short sentences to increase task difficulty. Results showed that overall accuracy by the Japanese group was significantly poorer than for the Americans (72% and 98% correct respectively). Certain clusters were harder for Japanese listeners (e .g., 76% correct for /pl/ but 64% for /bl/). Productions of the words were also obtained and are related to perceptual performance.

5pSWa9. Discrimination of Mandarin disyllabic tone contrasts by American English-speaking adults and children. Shari Berkowitz (CUNY Graduate Ctr., Speech Acoust. and Percept. Lab., 365 Fifth Ave., New York, NY, 10016, shariellen@gmail.com)

Most research on cross-language perception of Mandarin tones has used monosyllabic stimuli, despite the fact that Mandarin words are predominantly disyllabic. Previous work on disyllable discrimination (Berkowitz and Strange, 2007) suggested that perception by English-speaking adults is influenced by coarticulation and context effects. In this series of studies, subsets of stimuli were chosen to attempt to isolate the effects of fundamental frequency height, contour, turning point and timing on English speakers' ability to perform a same/different task. Three balanced blocks of approximately 10 min each were given, to keep the task appropriate for preschoolers. Feedback conditions varied; some subjects received feedback throughout and others on the first block only. Early results suggest that discrimination of disyllabic Mandarin stimuli is easier for American English speakers if the differences between disyllables occur at onset or offset of the complete utterance; differences in the middle pose more of a challenge. Subjects appear to evaluate disyllabic contours as a whole, rather than on a syllable-by-syllable basis. Implications for future research and training protocols will be discussed.

5pSWa10. Factors inducing cross-linguistic perception of illusory vowels. Jiwon Hwang (Dept. of Linguist., Stony Brook Univ., Stony Brook, NY 11794-4376, jihwang@ic.sunysb.edu)

Japanese speakers tended to hear an illusory vowel in illegal consonant sequences (Dupoux *et al.*, 1999). Korean has no sequences of stop followed by nasal; therefore, it is expected that Korean speakers would perceive an illusory vowel in stop-nasal. In an identification task comparing Korean and English listeners on stimuli along a continuum that ranged from no vowel (igna/ikna) to a full vowel (igVna/ikVna), Korean listeners reported the presence of a vowel significantly more often than English listeners, even when there is no vowel in the stimuli. However, this effect was found only when the stop was voiced, even though [kn] and [gn] are both illegal Korean sequences. In an AXB discrimination task, Korean participants had more difficulty discriminating 'stop-nasal' from 'stop-V-nasal' than English participants, again only when the stop was voiced. The results suggest that voicing, rather than simple illegality, induces bias toward perception of illusory vowel in Korean. This is explained by the fact that in Korean, voiced stops occur only prevocalically in intersonorant position (as an allophone of voiceless obstruents). This voicing effect is reflected in Korean L2 learners' production patterns in which they insert a vowel more often in voiced stop-nasal sequences than in voiceless stop-nasal sequences. [Work supported by NSF BCS-07460227.]

5pSWa11. Individual differences in the perception of final consonant voicing among native and non-native speakers of English. Bruce L. Smith (Dept. of Commun. Sci. and Disord., Univ. of Utah, 390 S. 1530 E., Rm. 1201, Salt Lake City, UT 84112, bruce.smith@hsc.utah.edu) and Rachel Hayes-Harb (Univ. of Utah, Salt Lake City, UT 84112)

Various studies have provided modest evidence supporting the ideas: (1) that non-native listeners outperform native listeners in judging productions by other non-native speakers with whom they share a first-language (L1) background and (2) that non-native listeners are more accurate judging productions by subjects with whom they share an L1 background than they are judging speech contrasts of native English speakers. Research regarding these issues has typically been based on group findings, so it is not clear to

what extent these tendencies may occur among individual listeners. The present investigation reports findings for 15 native English (NE) listeners and 15 non-native listeners (native Mandarin: NM), who made judgments about English voiced/voiceless minimal pairs produced by six other NE and six other NM talkers. When listening to the six NM talkers, fewer than half of the 15 NM subjects performed better than the range shown by the 15 NE listeners. Further, the 15 NM listeners were more accurate than the 15 NE listeners in judging the voicing contrast for just two of the six NM speakers. Thus, limited evidence was found that individual, non-native subjects demonstrate patterns that might have been expected on the basis of previous group findings.

5pSWa12. Perceptual confusions of American English vowels and consonants in Arabic-English bilinguals. Valeriy Shafiro (Dept. Comm. Disord. Sci. Rush Univ. Medical Ctr., 600 S. Paulina Str.,1015 AAC, Chicago, IL 60612), Anatoliy Kharkhurin (American Univ. of Sharjah, Sharjah, UAE), and Erika Levy (Columbia Univ., New York, NY 10027)

This study investigated the perception of American English (AE) vowels and consonants by proficient adult Arabic-English bilinguals studying in the United Arab Emirates (UAE). The native language of all participants was Arabic, and their average age of English acquisition was 6 years. In a closed set format, 29 participants were asked to identify 12 AE vowels presented in /hVd/ context and 20 AE consonants in three vocalic contexts: /ɑCɑ/, /iCi/, and /uCu/. Overall vowel identification was 70% correct. The three least-accurately identified vowels were /ɑ/, /ɔ/, /æ/, which were 17%, 38%, and 54% correct, respectively. Most confusions were found among low and back vowels. Overall, perception of consonants was more than 90% correct in all three vocalic contexts. The most errors on consonants were found for /ð/, i.e., 63% and 46% correct in /ɑCɑ/ and /iCi/ contexts, respectively. In both contexts, /ð/ was most frequently confused with /v/. However, identification of /ð/ in /uCu/ context was far more accurate; i.e., 97% correct. These results suggest a phonetic context effect on the cross-language perception of consonants.

5pSWa13. Experience with foreign accent influences non-native (L2) word recognition: The case of th-substitutions. Adriana Hanulikova and Andrea Weber (Max Planck Inst. for Psycholinguistics, Nijmegen, The Netherlands)

Effects of mispronunciations on word recognition are often explained in terms of perceptual similarity: the less similar mispronunciations are to target words, the more lexical activation is disturbed. Using th-mispronunciations that occur in foreign-accented English, this study investigated whether, irrespective of perceptual similarity, experience with mispronunciations influences word recognition. While Dutch speakers of English frequently substitute voiceless *th* with /t/ (e.g., /tɛft/ for *theft*), German speakers prefer /s/ (e.g., /sɛft/); the perceptually close /f/ occurs infrequently in both groups. Four eye-tracking experiments examined whether familiar substitutions cause stronger lexical activation than less familiar ones. German and Dutch participants listened to sentences spoken with a German or Dutch accent (e.g., "Now you will hear /tɛft/"), while they were looking at a display with four printed words (e.g., *theft*, *left*, *kiss*, *mask*). The time course of lexical activation was measured as a function of amount of looks to printed th-words after hearing mispronounced words with a /t/, /s/, or /f/ substitute. Irrespective of the heard accent, th-words were fixated more often when hearing /t/ for Dutch listeners but /s/ for German listeners, while /f/ never outperformed the accent-specific dominant substitute. The results suggest an influence of accent-specific experience on L2 word recognition.

5pSWa14. Cross-linguistic evidence for the influence of native language prosody in infant speech segmentation. Suzanne Curtin (Depts. of Psych. and Linguist., Univ. of Calgary, 2500 University Dr. NW, Calgary, AB T2N 1N4, Canada, scurtin@ucalary.ca), Linda Polka, Shani Abada (McGill Univ., Montreal, QC H3G 1A8, Canada), and Reaper Sally-Joy (Univ. of Alberta, Alberta, Canada)

Research using artificial languages with English-learning infants has shown both prosodic and distributional cues are used for speech segmentation by 7 months. When these cues conflict, infants younger than 7 months rely on distributional cues while older infants rely on prosodic cues). In the present study we assessed the role of prosodic information in segmentation when it is not favorably aligned with distributional cues, to determine whether language-specific rhythmic biases guide segmentation as suggested by studies using natural speech. Two continuous streams of naturally produced syllables (English and French) were constructed using nines syllables that are permissible in both languages. Within each stream statistical cues were manipulated independently of language-appropriate stress cues. English- and French-learning 8-month-olds were familiarized with their native language stream and then present test probe strings to determine what syllable sequences were extracted from the stream. Probes were selected to assess the role of stress cues in segmentation. Findings show that English infants make use of a trochaic template; Canadian-French infants show a weaker and less focused reliance on stress cues to segment words from connected speech. These cross-linguistic differences reflect processing biases that may be set by language experience and/or elicited by speech input properties.

5pSWa15. Individual gains from auditory training of English vowels for Greek native speakers. A. Lengeris (Dept. of Speech Hearing and Phonetic Sci., Univ. College London, 2 Wakefield St., London WC1N 1PF, UK, a.leggeris@ucl.ac.uk)

Several studies have shown that high-variability auditory training can improve the perception of second-language (L2) sounds by adult learners. However, even when testing a homogenous L2 group, considerable differences are commonly found between individuals not only in pre-training performance but also in how each trainee responds to training. Additionally, it is not clear what aspects of perceptual processing are actually being trained. To address these issues the present study trained Greek native speakers in perceiving the vowels of Southern British English. The trainees received five sessions of high-variability auditory training (including multiple words spoken by multiple talkers). The pre- and post-training tests included identification and discrimination tasks in L1 (Greek) and L2 (English), and a nonspeech (F2 only) discrimination task. Preliminary results show significant improvement in the trainees' identification of L2 vowels, but no significant change in their L2 discrimination. The relationships between pre-training L1, L2, and nonspeech performance and gains in L2 identification for individuals will be discussed.

5pSWa16. Vowel quantity and quality perception in Finnish and Estonian speakers. Stefan Werner (General Linguist. and Lang. Technol., Univ. of Joensuu, Finland, stefan.werner@joensuu.fi) and Einar Meister (Tallinn Univ. of Technol., Estonia)

Vowel quality perception in quantity languages might be expected to be unrelated to duration of the vowels since duration is used to realize phonemic duration oppositions. Our studies, both earlier reported ones and our latest project, indicate, though, that vowel-intrinsic duration changes have a consistent effect on vowel identification in Estonian and Finnish speakers. Our latest experimental setup for investigating perceptional effects of microduration consists of two blocks, first a set of tests to identify the subject's formant-based categorial boundaries between vowel pairs on the close-open axis, then the accordingly selected listening tests presenting formant-wise ambiguous vowel exemplars produced with varying durations. Results from ten Estonian and ten Finnish speakers will be presented. In addition, the phenomenon observed will be compared with formant structure and duration perception in languages where these parameters co-vary (e.g., English) and thus an influence of sub-phonemic durational changes on vowel perception would seem much likelier.

5pSWa17. The phonotactic influence on the perception of a consonant cluster /pt/ by native-English and native-Polish listeners: A behavioral and ERP study. Monica Wagner and Valerie L. Shafer (Ph.D. Prog. in Speech-Lang.-Hearing Sci. The Graduate Ctr., 365 Fifth Ave., New York, NY 10016-4309, mpwspl@aol.com)

The perception of the cluster /pt/ in word onset in native-Polish and native-English adult listeners was compared using behavioral and event-related potential (ERP) measures to explore the language-specific phonotactic influence on speech perception. Acoustic-phonetic features of phonemes vary with context and phonotactic rules specify all allowable phoneme contexts. Both native-Polish and native-English groups are exposed to the cluster /pt/ in their language but only one group, native-Polish, experience the cluster in the context examined, word onset. Word onset consonant cluster /st/ is used as a control. Participants were asked to determine whether

they heard the second word of a pair as consisting of two syllables or three (e.g., pteva versus peteva). Behavioral results revealed that Polish listeners perceive the /pt/ cluster with greater accuracy than English listeners. ERPs revealed late components that differ for the two language groups reflecting linguistically-relevant aspects of the /pt/ contrast and early components that are essentially the same in the two language groups reflecting physical-acoustic differences. ERP results suggest the /st/ contrast to be more salient than the /pt/ contrast consistent with the higher behavioral accuracy for the /st/ cluster. These findings support the suggestion that psychoacoustic salience of acoustic-phonetic features influences speech perception.

5pSWa18. The roles of second-language experience and accent in speech recognition in noise. Melanie Preece-Pinet and Paul Iverson (UCL, Dept. of Lang. Sci., Chandler House, 2 Wakefield St., London WC1N 1PF, UK, m.pinet@ucl.ac.uk)

Previous work has demonstrated that there is an interaction between native (L1) and non-native (L2) accents in speech recognition in noise, with listeners being better at L1 or L2 accents that match their own speech. This study investigated how L2 experience modulates this talker-listener interaction. L1 southern British English (SE) and L1 French listeners with varying L2 English experience (inexperienced FI, experienced FE, and bilinguals FB) were tested on the recognition of English sentences mixed in speech-shaped noise that was spoken with a range of accents (SE, FE, FI, Northern Irish, and Korean-accented English). The results demonstrated that FI listeners were more accurate with strongly accented FI talkers, and were progressively worse for the other accents, perhaps based on accent similarity. The SE listeners, however, had a strong advantage for SE speech, but were similarly poor at understanding the other accents. Their recognition processes were thus selectively tuned to their own accent, rather than having the graded sensitivity of FI listeners. FE and FB listeners were more similar to SE listeners as their experience with English increased. The results thus suggest that increased exposure to L1 speech changes both the intelligibility of accents and the selectivity of accent processing.

5pSWa19. Second language word segmentation in a fluent speech listening task. Tuuli M. Adams (Dept. of Linguist., New York Univ., 726 Broadway, New York, NY 10003, tuuli.adams@nyu.edu)

Extracting words from a continuous speech stream is a complex task involving the integration of multiple linguistic cues. Like infants, adults have been shown to use both distributional information through statistical learning, and metrical and intonational information, when segmenting words in a second language. This study investigates adults' native language influence on speech segmentation, and the extent to which linguistic knowledge is augmented by exposure to fluent, natural speech in a second language. In the experiment, English listeners completed a word learning task in an unfamiliar language, Finnish. Next, they listened to a fluent speech passage which contained learned words, as well as syllable sequences controlled for the same frequencies and transitional probabilities as the real words. English speakers have been shown to use acoustic stress cues to locate word boundaries, and while Finnish has a regular stress pattern, other phonotactic patterns could affect English listeners' segmentation abilities. Afterwards, they completed a forced-choice identification task, choosing from pairs of real and non-words, including incorrectly segmented syllable sequences, to identify the token which is more likely a word in Finnish. The behavioral data are analyzed for patterns reflecting the integration of native language linguistic knowledge and statistical learning abilities.

5pSWa20. Assessing the contribution of second language experience and age of learning in Catalan/Spanish bilinguals' perrception and production of English sounds. Natalia Fullana and Ian R. A. MacKay (Dept. of Linguistcs, Univ. of Ottawa, 70 Laurier Ave. East, Ottawa ON, K1N 6N5 Canada, fullnat@gmail.com)

Previous research in immersion settings has shown that an early age of onset of second language (L2) learning, together with long-term exposure to the L2, are determinant factors for perceiving and producing L2 sounds accurately [e.g., Flege, MacKay, & Meador (1999)]. However, research in formal learning contexts has resulted in negative evidence for an early age of learning advantage [e.g., Garcia-Lecumberri & Gallardo (2003)] or in divergent experience effects (Cebrian, 2003, 2006). This study aimed to further examine the contribution of the factors of age of onset of L2 learning (AOL) and experience in a foreign language learning environment. Catalan/Spanish bilinguals studying English at university, with AOLs of 4 to 14 years and a minimum of 7 years of formal instruction, performed an AXB discrimination task, a picture narrative, and a delayed sentence repetition task. Results revealed that Catalan/Spanish bilinguals with somewhat longer exposure to English and an earlier AOL tended to discern English sounds at higher correct rates. By contrast, a great degree of variability was found across the bilinguals' extemporaneous and prompted production of English segments. Findings are discussed in terms of current models of L2 speech acquisition and their application to formal learning settings. [Work supported by postdoctoral fellowship from the Ministerio de Educacion y Ciencia and the FECYT (Spain).]

5pSWa21. Asymmetric development of perception and production of lexical stress in Korean second-language (L2) learners of English. Jeonghwa Shin and Shari/R. Speer (Dept. of Linguist., Ohio State Univ., 1712 Neil Ave., Columbus, Ohio 43210, jshin@ling.osu.edu)

This study explores perception and production of lexical stress information in L2 English learners whose L1 employs a fixed rhythmic pattern at the lexical level. Nineteen English L1 speakers and 14 Korean L2 learners of English were trained to learn 16 minimal stress nonword pairs with picture referents which are segmentally disambiguating in the last syllable (/dʒákunaɪ/ vs /dʒakúnəɹ/). The eye-tracking perception experiment revealed that English L1 speakers exploited lexical stress information of the first two syllables to spot the target word in the instruction, "Click on the (target word)," whereas Korean L2 learners' identification of the target word was delayed until the last syllable. In their production of words in a carrier, "This is the (target word)," however, L2 learners used loudness and durational cue to indicate alternative prominence of the first two syllable as did English L1 speakers. These results imply the development of production and processing of lexical stress in L2 word learning can be asymmetrical. This asymmetry is attributable to the L1 which has a fixed (T)HLH accentual phrase pattern at the lexical level and thus provides little incentive for the lexically specified stress processing. [Work supported by Linguistics Dept., Ohio State University.]

5pSWa22. Do first-language (L1) phonemic categories play a role in the perception of second-language (L2) phonemic contrasts? A look from the perception of Cantonese codas by Mandarin speakers. Patrick Chun Kau Chu (School of Psych., The Univ. of New South Wales, Sydney, Australia and Dept. of Linguist. and Modern Lang., The Chinese Univ. of Hong Kong, Hong Kong, China, patrickhk83@hotmail.com)

This study investigated whether the phonological system of L1 or universal hierarchy of perceptual similarities has a larger effect on Mandarin speakers' perception of Cantonese phonemic categories. In Mandarin and Cantonese, there are differences in the distribution of nasals and plosives in the coda position. Both nasals (/m/, /n/, and /ng/) and plosives (/p/, /t/, and /k/) are possible codas in Cantonese while only nasals (/n/ and /ng/) can appear in the coda position in Mandarin. Non-native phonemic categories (e .g., /t/) will be assimilated to a native phonemic category (e.g., /n/) if L1 plays a larger role while assimilation to a non-native phonemic category (e .g., /p/ or /k/) may support the idea that universal hierarchy of perceptual similarities plays a more important role. An AX discrimination task is carried out to see whether Mandarin and Cantonese speakers can discriminate these six phonemic contrasts in Cantonese monosyllabic words. Results showed that Mandarin speakers have poorer discrimination ability in the later type (e.g., /t/ vs /k/) than the former type (e.g., /t/ vs /n/). This suggests that the Perceptual Assimilation Model may need to take universal hierarchies of perceptual similarities into account when predicting L2 learners' discrimination ability in a second language.

5pSWa23. Implicit learning of non-native speech stimuli. Eleni L. Vlahou, Athanassios Protopapas (ILSP/Athena, Artemidos 6 & Epidavrou, GR-151 25 Maroussi, Greece, evlahou@ilsp.gr), and Aaron Seitz (Univ. of California-Riverside, Riverside, CA 92521)

Previous studies have induced robust adult learning of non-native natural speech stimuli using explicit instructions and feedback. Here, Greek adults were exposed to non-native Hindi phonemes for six sessions in which they were unaware of the phoneme distinctions and the phonemes had no relevance to their main task. Stimuli were natural recordings of CV syllables (a

retroflex or dental unvoiced stop followed by [a]) from a Hindi speaker. On each trial, participants heard two identical dental sounds and two retroflex sounds differing in intensity by an adaptively varied amount between 0.5–3 dB. The explicit task was to identify which pair differed in intensity, while the implicit learning goal was the phonetic contrast. After training, participants were tested on explicit phonetic identification and discrimination of trained and untrained stimuli spoken by the same and a novel speaker. The trained group outperformed an untrained group who were administered the same tests. Our results are compatible with a perceptual learning model where task-irrelevant stimuli (Hindi sounds) are learned due to a systematic pairing with task targets (intensity difference). Learning did not extend to the novel speaker, in agreement with previous studies showing that variability in training is crucial for generalization.

5pSWa24. Duration as a phonetic cue in the categorization of /i:/-/ɪ/ and /s/-/z/ by spanish/Catalan learners of English. Eva Cerviño and Joan C. Mora (Dept. Filologia Anglesa, Universitat de Barcelona, Gran Via 585, 08007-Barcelona, Spain, ecervino@ub.edu)

Previous research has shown that Spanish/Catalan learners of English (SCLs) rely on duration rather than spectral cues in the categorization of contrastive English vowel pairs such as /i:-/ɪ/, despite length being nondistinctive in Spanish/Catalan. Native English speakers (NESs) rely more heavily on V-C duration ratio than on glottal pulsing in the perception of voicing contrasts in -VC contexts. This study investigated whether SCLs' overreliance on duration in vowel categorization also applied to consonants in contexts (-VC) where lack of closure voicing makes duration the main cue to voicing for NESs. SCLs ($N=40$) and NESs ($N=15$) participated in two lexical decision tasks consisting of /f/+V+/t/ stimuli drawn from two 8-step continua with shortened (/i:/) and lengthened (/ɪ/) vowels; and /p/+/i:/+C stimuli drawn from six 8-step continua with simultaneously and independently modified vowels and final consonants (/i:s/-/i:z/). Manipulating duration had a significant effect on SCLs' categorization of /i:/ and /ɪ/, but not on SCLs' categorization of /s/ and /z/, suggesting that SCLs made use of duration as a cue to vowel categorization, but did not use duration as a cue to voicing. These results are interpreted and further discussed in light of Bohn's Desensitization Hypothesis and McAllister's Feature Hypothesis.

5pSWa25. Phonetic cue-weighting in the acquisition of a second language: Evidence from Greek speakers of English. Anastasia Giannakopoulou, Maria Uther (Ctr. for Cognition and Neuroimaging, Dept. of Psychol., Brunel Univ., Uxbridge, Middlesex, UB8 3PH, UK), and Sari Ylinen (Univ. of Helsinki, PL 9, Helsinki, 00014, Finland)

The most difficult speech-sound contrasts for foreign-language learners are often the ones that have multiple phonetic cues, especially if the cues are weighted differently in the foreign and native languages. Greek speakers of English were studied to determine on what basis they are making discrimination and identification between English vowels. The use of spectral and duration cues of English vowel minimal pairs (e.g., "beat" versus "bit") in Greek speakers were studies using perception and discrimination tasks. There were two forms of perceptual, minimal pair identification tasks. One used natural stimuli and the other used matched vowel duration words to "force" the use of spectral cues. Performance was impaired for Greek speakers across both perceptual identification tasks compared to native speakers. But performance was the worst for Greek speakers for the matched duration condition. Interestingly, an AB-X discrimination task with the same stimuli did not hamper the performance of the Greeks compared to the English in the same way. Pilot data from Greek children also showed similar patterns of performance across conditions to the Greek adults, but with a somewhat reduced ceiling. These results are discussed in terms of the strategies used to acquire new languages.

5pSWa26. Role of rhythmic and intonational cues in language discrimination. Chad Vicenik and Megha Sundara (Dept. of Linguist., 3125 Campbell Hall, UCLA, Los Angeles, CA 90095)

Languages differ in rhythm as well as intonation. Research indicates that adult listeners are able to use rhythm to discriminate between two languages from different rhythm classes [Ramus Mehler, (1999)]. For languages within the same rhythm class, adults are able to use to use intonation to discriminate between languages like English and Dutch, but only when one of the languages is familiar [Ramus and Mehler, (1999); Pijper, (1983)]. It remains unclear if the rhythmic differences between languages in the same rhythm class are enough to support language discrimination. In this paper, we tested American English listeners' ability to categorize re-synthesized American English and German sentences or American and Australian English sentences from which all segmental information had been removed. English and German are from the same rhythm class and differ in intonation; whereas American and Australian English can be thought to be rhythmically identical, but differ in intonation. Subjects were tested in three conditions: (a) with only intonational cues; (b) with only rhythmic cues; and (c) both intonational and rhythmic cues. Preliminary results indicate that listeners rely on rhythm, but not intonation, to distinguish between English and German, and only on intonation to distinguish between American and Australian English.

5pSWa27. Discrimination of American vowels in disyllables mixed in speech babble by experienced Japanese and Russian learners of English. Kikuyo Ito, Yana, D. Gilichinskaya, and Winifred Strange (Speech Acoust. and Percept. Lab., CUNY-Graduate Ctr., 365 Fifth Ave., New York, NY, 10016-4309, kikuyoito@hotmail.com)

Previous studies in this laboratory examining speeded discrimination of American English (AE) vowels in quiet by Japanese (JP) and Russian (RU) late L2 learners indicated differences in relative difficulty of non-native contrasts that were predictable from L1 phonological differences. Response latencies were a sensitive measure of continuing L2 perceptual difficulties. In the present study, the same task (a speeded categorial ABX task with disyllabic stimuli) was administered to Japanese and Russian listeners, with stimuli mixed in speech babble at three levels (SNR 0, 6, 12). Eight experimental contrasts (four with spectral plus duration differences and four with spectral differences only) and four control contrasts were tested in lists with both vowel contrast and noise level varying randomly. Results indicated that both L2 groups performed more poorly than native AE control subjects on many of the experimental contrasts, including those differentiated by duration as well as spectral differences. Again differences in relative perceptual difficulty across L2 groups emerged. Perceptual assimilation patterns and performance on a test of English fluency and pronunciation (versant) were correlated with discrimination performance across individuals. [Work supported by NSF.]

5pSWa28. Phonetic accommodation in conversations between native and non-native speakers. Midam Kim (Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, midamkim@gmail.com)

We explored phonetic accommodation during native speaker interactions (four native English pairs and four native Korean pairs) and native–non-native interactions (four native–non-native pairs, all speaking English). Speakers engaged in a collaborative picture description task that lasted approximately 20 min. Utterances from relatively early and late in the task were subsequently presented to native English listeners (or Korean listeners, for the Korean-language dialogues), who were asked to carry out *XAB* similarity judgments, where X=one talker's utterance and A/B=early and late utterances from the partner. Phonetic accommodation is indexed by how frequently listeners select the "late" utterance as sounding more similar to the target utterance. An independent group of native English speakers rated the nonnative utterances for degree of accentedness. Phonetic convergence was observed for native-native conversations (English or Korean), and this was stronger when speakers shared the same or similar dialects. In native–non-native conversations, none of the native speakers converged towards a non-native partner, while non-native speakers showed different patterns depending on their proficiency, with greatest convergence for moderately accented nonnatives. The results suggest that phonetic accommodation can occur cross-linguistically, and that it may be constrained both by speakers' dialect and by their language proficiency.

5pSWa29. Speeded discrimination of American English vowel contrasts by Spanish-speaking late second-language (L2) learners. Jason Rosas, Arsenia Barias, Yana D. Gilichinskaya, and Winifred Strange (Speech Acoust. and Percept. Lab., CUNY-Graduate Ctr., 365 Fifth Ave., New York, NY 10016-4309)

This study is the third in a series exploring discrimination of American English (AE) vowel contrasts by late L2-learners of English whose native languages have small vowel inventories (Japanese, Russian, Spanish). On a speeded ABX task, adult Spanish-speaking listeners discriminated multiple

AE vowel contrasts including adjacent height pairs that differed in both spectral and duration parameters, front/back pairs (differing only in spectral characteristics, and four nonadjacent control pairs). It was hypothesized that Spanish listeners, who do not distinguish native vowels based on duration, would have difficulties with both types of experimental contrasts, relative to control pairs. Additionally, mean reaction time (RT) difference scores (relative to mean RTs for control pairs) for Spanish listeners were expected to be slower than for AE controls. Preliminary findings confirm these predictions. Differences in relative difficulty of experimental pairs compared with earlier data for Japanese and Russian L2-learners showed L1-specific patterns of performance that were related to L1/L2 perceptual similarity patterns. Spanish subjects also completed the Versant Test, an English language proficiency examination conducted by phone. A correlational analysis of discrimination accuracy and speed with Versant Test fluency and pronunciation measures will be presented. [Work supported by NSF.]

5pSWa30. Spoken word recognition in quiet and noise by native and non-native listeners: Effects of age of immersion and vocabulary size. Astrid Z. Doty, Catherine L. Rogers (Dept. of Comm. Sci. and Dis. and Psych., Univ. of South Florida, 4202 E. Fowler Ave. PCD1017, Tampa, FL 33620), and Judith B. Bryant (Univ. of South Florida, Tampa, FL 33620)

In spoken word recognition, high-frequency words with few and less frequently occurring minimal-pair "neighbors" (lexically easy words) are recognized more accurately than low-frequency words with many and more frequently occurring neighbors (lexically hard words). [Bradlow and Pisoni, J. Acoust. Soc. Am.,106, 2074–2085 (1999)] found a larger "easy-hard" word effect for non-native than native speakers of English. The present study extends this work by specifically comparing word recognition by non-native listeners with either earlier (age 10 or earlier) or later (age 14 or later) ages of immersion in an English-speaking environment to that of native English speakers. Listeners heard six lists of 24 words, each composed of 12 lexically easy and 12 lexically hard target words in an open-set word-identification task. Word lists were presented in quiet and in moderate background noise. A substantially larger easy-hard word effect was obtained only for the later learners, but a measure of oral vocabulary size was significantly correlated with performance for the non-native listener groups only. Thus, the increased easy-hard effect for non-native listeners appears to be better explained as an effect of phonetic proficiency and/or the effect of vocabulary size on the structure of lexical neighborhoods than as an effect of language competition.

5pSWa31. Perceptual adaptation to foreign accented speech. Melissa Baese-Berk (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, m-baese@northwestern.edu)

Previous research on native listener adaptation to foreign-accented adaptation has demonstrated that training on a single non-native speaker results in talker-dependent learning (i.e., training does not generalize to new talkers). However, training on multiple foreign-accented speakers from a single language background results in talker-independent, but accent-dependent learning (i.e., training generalizes to new speakers of the trained accent, but not to new accents). The current study extends these findings, examining whether training on multiple accents can result in accent-independent learning. Native English listeners were trained on recognition of foreign-accented speech by five speakers from different language backgrounds. They were then tested on the speech of two novel speakers. One of the speakers was a native speaker of a language included in the training set (Mandarin), and the other was a native speaker of a language not included in the training set (Slovakian). Listeners demonstrated better performance on both the Mandarin and Slovakian tests than untrained control subjects, demonstrating accent-independent learning after training on multiple foreign accents. This suggests that individuals in multilingual communities, who are exposed to significant variation in the input, may be able to achieve highly flexible speech perception systems.

5pSWa32. Effects of audiovisual auditory and articulatory training on second-language (L2) vowel category learning. Cristina Aliaga-Garcia (Departamento de Filologia Inglesa y Alemana, Universitat de Barcelona, Gran Via Corts Catalanes 585, 08007 Barcelona, Spain, cristinaaliaga@ub.edu)

Cross-language speech perception research has documented learners' difficulties in using and accessing L2 durational and spectral information. This study investigated the short-term effects of two types of audiovisual high-variability phonetic training—auditory versus articulatory training—on the perception of the 11 English RP monophthongal vowels. Bilingual Catalan /Spanish learners of English ($N=20$ in each training condition) and one control group ($N=20$) were tested before and after training. Perceptual gains were assessed through a four-choice categorization task with natural CVC words, a forced-choice categorization task based on four synthesized /hVd/ continua (/i:/-/ɪ/, /æ/-/ʌ/, /ʌ/-/ɑ:/ and /u:/-/ʊ/), and an AX discrimination task including 16 contrasting pairs. Significant perceptual accuracy gains were obtained in both auditory and audiovisual conditions. Pre- and postcategorization of synthesized vowels suggests that vowel perception improved, and that both training modalities had a similar positive effect on the learners' use of spectral and duration information in vowel categorization. These results are further discussed in terms of input effects on vowel category learning and the reweighting of acoustic cues through phonetic training.

5pSWa33. Semantic contextual cues and listener adaptation to foreign-accented English. Page Piccinini (2146 Sherman Ave., Apt. 1D, Evanston, IL 60201, p-piccinini@u.northwestern.edu) and Ann Bradlow (Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, abradlow@northwestern.edu)

The context surrounding words can facilitate word recognition for native and non-native listeners. However, for degraded speech signals, (e.g., with noise), non-natives may lose the ability to take advantage of contextual cues unless the speech is clearly produced [Bradlow and Alexander (2007)]. This study investigated whether: (1) foreign-accented speech degrades the signal such that natives show reduced ability to take advantage of context, and (2) both native and non-native listeners adapt to foreign-accented speech. Native and non-native listeners were exposed to accented speech in two blocks. In each block, half the sentence final words were in high and half were in low predictability contexts. Listeners were asked to identify sentence final words. In a single talker condition, natives benefited from context and adapted to the accent; non-natives could not use context but showed adaption. In a multitalker condition, natives neither benefited from context nor adapted to the accent. Non-natives benefited from context and adapted; however, the context effect was only seen when listener and speaker L1 matched. These results suggest that accented speech disrupts the native ability to take advantage of context, but this can be overcome by adaptation to the speaker. Non-natives benefit substantially from a talker-listener L1 match.

5pSWa34. Perception of dialectal variation: Can speakers of Western Canadian English perceive New Zealand English /r/-sandhi? Verona J. Dickout and Benjamin V. Tucker (Dept. of Linguist., Univ. of Alberta, 4-32 Assiniboia Hall, Edmonton, AB, Canada T6G 2E7)

The current study investigates perception of New Zealand English /ɾ/-sandhi by speakers of Western Canadian English. It investigates whether speakers of Western Canadian English can differentiate between intrusive-/r/ (e.g., *sawing* [sæoɹəŋ]) and linking-/r/ (e.g., *soaring* [sæoɹəŋ]) of /ɾ/-sandhi across morpheme boundaries and across word boundaries (e.g., linking: *boar and*, intrusive: *boa and*). Reaction time and accuracy were recorded in a two-alternative forced choice experiment. Stimuli showed significant acoustic differences between the linking-/r/ and intrusive-/r/ environments, with intrusive-/r/ having a shorter duration than linking-/r/. Participants were highly accurate and faster at recognizing words with linking-/r/ (e.g., *soaring*) and extremely inaccurate and slower at recognizing intrusive-/r/ (e.g., *sawing*). Participants' responses to linking-/r/ at a morpheme boundary were significantly more accurate and faster than responses with linking-/r/ across word boundaries. The results of this experiment address the phonetic variability present in cross-dialect perception and begin to investigate strategies listeners utilize to accurately perceive speech.

5pSWa35. Perception of phonemic and allophonic contrasts in English- and Spanish-learning infants. Morgan Meredith (Commun. Sci. & Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208, morgan.leigh@me.com) and Jessica Maye (Northwestern Univ., Evanston, IL 60208)

Infants from monolingual English- and Spanish-speaking homes were compared in their perception of two phonetic contrasts that differ in their phonemic status in the two languages. The contrast [d]-[ð] is phonemic in English but allophonic in Spanish; while the contrast [d]-flap is phonemic in Spanish but allophonic in English. Infants ranged from 6 to 16 months of age. Each infant was tested on discrimination of both contrasts. For half of the infants the stimuli were produced by a Spanish speaker, while for the other half of the infants the stimuli were produced by an English speaker. Data collection from Spanish-learning infants is still in progress. However, preliminary results indicate that infants in the youngest age group (6–8 months) are able to discriminate the [d]-flap contrast regardless of infant language background, while the [d]-[ð] contrast is discriminated poorly at this age in both language groups. English-learning infants lose sensitivity to the allophonic [d]-flap contrast with age, and it is not until 14–16 months that discrimination of the phonemic [d]-[ð] contrast improves in these infants. Spanish-learning infants are expected to remain proficient at discriminating the phonemic [d]-flap as they age and to remain poor at discriminating the allophonic [d]-[ð] contrast. [Work supported by NIH1R03HD048538 to JM.]

5pSWa36. The effects of linguistic experience on the perception of breathy phonation. Christina Esposito (Dept. of Linguist., Macalester College, 1600 Grand Ave., St. Paul, MN 55105, esposito@macalester.edu)

This study investigates the role linguistic experience has on the perception of phonation and the acoustic properties that correlate with this perception. Listeners from Gujarati (contrasts breathy versus modal vowels), Spanish (no breathiness), and English (allophonic breathiness) participated in two tasks. In the visual sort, subjects sorted breathy and modal vowels from a variety of languages into two groups based on perceived similarity of the talkers' voices. In the multidimensional scaling task, listeners heard pairs of Mazatec vowels, and moved a slider to indicate perceived similarity of vowels in each pair. Results showed that judgments were more uniform across subjects who had breathy categories present in their native language and varied across subjects who lack a breathy category. Despite allophonic breathiness, English listeners did no better at distinguishing phonations than Spanish listeners. The perceived similarity among the stimuli correlated with measurable acoustic properties (H1-H2, H1-A3, H1-A1, etc.). H1-H2 was the most salient acoustic property for Gujarati listeners (which correlates with their production of breathy vowels). English listeners relied weakly on the measure H1-H2, which is associated with the production of phonation in English. In addition, Spanish listeners, despite lacking a breathy category, relied on H1-H2 and H1-A1.

5pSWa37. Native language experience influences the perceived similarity of second language vowel categories. Thomas A. Farmer, Ran Liu, Neha S. Mehta, and Jason D. Zevin (Dept. of Psychiatry, Sackler Inst. for Develop. Psychobiol., Weill-Cornell Medical College, Box 140, New York, NY 10021, taf22@cornell.edu)

Most studies of L2 speech perception seek to characterize—at least implicitly—how the similarity among L2 speech sound categories is shaped by L1 experience. Intercategory similarity for native language speakers is rarely considered, however. Here, we derive two indices of graded intercategory similarity for a front vowel series (pin, pen, and pan): One from an off-line perceptual similarity judgment task and a second from online measures of arm-movement trajectories in a word recognition task. Both tasks revealed graded effects of intercategory similarity, but the similarity spaces differed between language groups. Both groups perceived /ae/ and /E/ to be most similar, but the native Italian speakers perceived /I/ to be equally similar to /E/ and /ae/, whereas English speakers perceived the /I/ and /E/ to be more similar than /I/ and /ae/. The Italian speakers' performance in both tasks suggests a similarity space that is dominated by dimensions that are contrastive for /I/ with respect to the other vowels. This may reflect a compensatory strategy for L2 vowel perception that may derive from experience with the distributional properties of their native vowels.

5pSWa38. The role of linguistic experience in the hemispheric perception of Mandarin tone. Shuang Lu (Prog. in Linguist., Univ. of Florida, Gainesville, FL 32611-5454, shuanglu@ufl.edu), Vincent J. van Heuven (Leiden Univ., 2300 RA Leiden, the Netherlands), and Ratree Wayland (Univ. of Florida, Gainesville, FL 32611-5454)

The current study investigated whether and to what extent lexical tone is processed as linguistic information by native Mandarin listeners, early Mandarin-Dutch bilinguals, second year Dutch students of Chinese, and monolingual Dutch students. Dichotic listening task was conducted in each of the four groups in order to determine to what extent the processing of lexical tones in Mandarin was lateralized to the left hemisphere. The majority of native Mandarin listeners revealed a right-ear advantage (REA) corresponding to a left-hemisphere lateralization in the perception of Mandarin tones. Similar to the native Chinese listeners, the Dutch-Mandarin bilinguals also exhibited an REA. Moreover, the left-hemisphere dominance was even more evident for these proficient bilinguals than for the native listeners. However, the Dutch listeners without any experience with Mandarin or other tonal languages showed no ear advantage. For the second year Chinese major listeners, five listeners revealed an REA, while others exhibited a bilateral pattern. The results suggested that there is a tendency for a right-ear advantage corresponding to left-hemisphere superiority for listeners with at least some basic experience of Mandarin tones.

5pSWa39. Category formation and lexical encoding of a new contrast. Isabelle Darcy, Laurent Dekydtspotter, Rex A. Sprouse (Dept. of Second Lang. Studies, Indiana Univ., 1021 E Third St., Bloomington, IN 47405, idarcy@indiana.edu), Christiane Kaden, John H. G. Scott, Vance Schaefer, and Michael McGuire (Indiana Univ., Bloomington, IN 47405)

The question whether category formation is a prerequisite for U.S.-English learners of French to encode a non-native contrast in lexical representations is investigated, looking at front [y-œ] and back [u-ɔ] rounded vowels. An ABX categorization experiment revealed no group difference between advanced ($N=18$) and inexperienced learners ($N=18$). Both made significantly more errors than French controls ($p<0.01$) on the [u-y] contrast, despite a good global discrimination (15% error). The possibility that minimal pairs of difficult contrasts (e.g., *sourd* [sur] deaf vs *sur* [syr] sure) are stored as homophones was tested in a lexical decision task with repetition priming. French words and non-words were paired with either themselves (*repetition*) or a minimal-pair-counterpart (*minimal pair*) in a 260 item list. Correct RTs were measured for each item. Given a comparable RT-advantage on the *repetition* and the *minimal pair* condition, merged lexical representations were assumed. Advanced learners, like native speakers, showed no RT-advantage for *minimal pair* conditions; inexperienced learners displayed significant facilitation for [u-y] and [ɔ-œ] minimal pairs (but not [i-y] control condition). This suggests that successful lexical contrast is possible for advanced English-French users despite persistent perception errors—the hallmark of an insecure category establishment—presenting an argument for the dissociation of both mechanisms.

5pSWa40. Relationship of prosody by Spanish speakers of English as a second language on the perception of intelligibility and accentedness by native English listeners. Paul Edmunds (Dept. of Linguist., Univ. of New Mexico, MSC06 3850, 1 University of New Mexico, Albuquerque, NM 87131)

The degree to which prosodic features related to English lexical stress affect the perception of factors such as intelligibility and accentedness by native English listeners was investigated. Acoustic analyses of English multi-syllabic words spoken by native Spanish speakers who learned English as a second language (ESL) were used to determine the values of the acoustic correlates of lexical stress of these speakers on a set of target words. Values for vowel duration, intensity, and fundamental frequency produced by the Spanish ESL speakers were compared to those values produced by native English speakers. Prosodic variation observed in the ESL speakers was used to formulate a range for manipulations to the target words. To focus solely on prosody and not take into account consonant and vowel quality, manipulations, using the PSOLA algorithm in Praat, were done to the

voices of native English speakers. These manipulations captured increases and decreases in duration, intensity, and fundamental frequency on vowels that should or should not carry lexical stress in a particular target word. Listeners rated the resynthesized tokens for factors such as intelligibility and accentedness, and the results suggest that a speaker's prosody alone can influence a native listener's judgments on these variables.

5pSWa41. The benefit of variation in cross-language perception of voice onset time (VOT). Meghan Sumner (Dept. of Linguist., Margaret Jacks Hall, Bldg. 460, Stanford Univ., Stanford, CA 94305-2150)

Listeners are consistently able to adjust to perceptually ambiguous sounds (Norris *et al.*, 2003; among others). When listeners are exposed to sounds ambiguous between two phonetic categories, boundaries adjust to the ambiguous signal. Perceptual learning has been suggested to account for adjustment to unfamiliar accents. It is not necessarily true, though, that this flexibility extends to more extreme cases of variation. Many sounds in accented speech are unambiguously mapped to the wrong category (e.g., unaspirated voiceless stops in French-accented English are perceived as voiced by native English listeners). We examined English listeners' perception of accented English in a perceptual learning paradigm with two exposure conditions: native French VOT mean and range in words that are lexically unambiguous (e.g., poach, *boach, not pet bet). Perceptual learning completely fails when listeners are exposed to the mean VOT, and in fact, this is coupled with an expanded voiced category, even though the words are lexically biased. Listeners make large adjustments when exposed to a range of VOTS, even those that are not good exemplars of English voiceless stops. The results are counter to models that suggest less variability improves learning, and suggest that variability in speech is used by and beneficial for listeners.

5pSWa42. The impact of native inventory versus native contrast on vowel discrimination. Marc Ettlinger (The Roxelyn and Richard Pepper Dept. of Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208, marc@northwestern.edu) and Keith Johnson (Univ. of California, Berkeley, Berkeley, CA 94720)

Theories of speech perception differ as to whether experience with sounds or with sound contrasts is more important when discriminating novel sounds. To explore this question, 21 American English, 16 Turkish, and 16 French speakers were tested on their ability to discriminate high front vowels in German (/i/~/y/~/Y/~/I/). The crucial difference between the participant languages is that English only uses the tense-lax contrast (/i/~/I/), while Turkish and French only use the rounding contrast (/i/~/y/). If contrast is crucial, then English speakers should be better at discriminating /y/~/Y/, whereas if the inventory is crucial, then they should be better at discriminating /I/~/Y/. The results of a fixed discrimination task show that there is no significant effect of language and that the tense-lax contrast (/i/~/I/; /y/~/Y/) is universally more easily discriminated than the rounding contrast (/i/~/y/; /I/~/Y/). The results of a rating task do reflect an influence of language, however with English-listeners rating /I/~/Y/ as more distinct and French and Turkish speakers rating /y/~/Y/ as more distinct. So, because English has /I/, but no rounding contrast, what matters in the perception of new sounds is not experience with contrasts, but rather experience with particular sounds.

FRIDAY AFTERNOON, 22 MAY 2009 SKY BRIDGE AUDITORIUM, 3:15 TO 4:45 P.M.

Session 5pSWb

Speech Workshop: Training and Adaptation of Speech Perception in Laboratory Contexts

Paul Iverson, Chair
Phonetics and Linguistics, University College London, London NW1 2HE, U.K

Contributed Papers

3:15

5pSWb1. Experience with computerized speech-perception training (SPATS-ESL) for speakers of other languages learning English. James D. Miller, Roy Sillings, Charles S. Watson (Commun. Disord. Technol., Inc., 501 N. Morton Str., Sta. 215, Bloomington, IN 47404), Isabelle Darcy, and Kathleen Bardovi-Harlig (Indiana Univ., Bloomington, IN 47405)

SPATS-ESL evolved as it was used by volunteers and as a supplement to classes in the Intensive English and English Enhancement Programs at Indiana University (Bloomington). These trials with eighty ESL-learners representing 12 L1s resulted in the curriculum described in the companion poster. Before training, the ESL learners exhibited significant problems with perceiving spoken English. These were greatest for syllable nuclei, least for syllable onsets, and intermediate for syllable codas. Recognition of spoken sentences was also inferior to that of native speakers of English. Improvement on all tasks including sentence recognition was proportional to the total time spent with SPATS. Based on the data and on reasonable extrapolations therefrom, many ESL-learners, after 20–30 hours of spaced practice with SPATS-ESL, appear to be approaching the performance of native-speakers of English on all SPATS-ESL tasks. Responses to questionnaires and comments indicate that the participants believed that SPATS-ESL training (a) was useful, (b) should be used in other classes, (c) helped them understand native speakers in real-life situations, and (d)would help them to improve their pronunciation of English. More advanced students were more enthusiastic about SPATS-ESL than less advanced students. Individual and group data supporting these generalizations and extrapolations will be presented.

3:33

5pSWb2. The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. Maria V. Kondaurova (Dept. of Otolaryngol., Head and Neck Surgery, Indiana Univ. Sch. of Medicine, 699 West Dr., RR044, Indianapolis, IN 46202, mkondaur@iupui.edu) and Alexander L. Francis (Purdue Univ., West Lafayette, IN 47906)

This study investigates the role of two processes, cue enhancement (learning to attend to acoustic cues which characterize a speech contrast for native listeners) and cue inhibition (learning to ignore cues that do not), in the acquisition of the American English tense and lax ([i] and [ɪ]) vowels by native Spanish listeners. This contrast is acoustically distinguished by both vowel spectrum and duration. However, while native English listeners rely primarily on spectrum, inexperienced Spanish listeners tend to rely exclusively on duration. Twenty-nine native Spanish listeners, initially reliant on vowel duration, received either enhancement training, inhibition training, or training with a natural cue distribution. Identification results demonstrated that, although listeners in all training groups increased their reliance on spectrum while decreasing reliance on duration, inhibitory and enhancement methods were significantly better than natural distribution training. Adaptive training was also more successful than the other methods for increasing be-

tween category distinctiveness on a discrimination task. These results suggest that phonetic learning may involve two distinct cognitive processes, cue enhancement and cue inhibition, that function to shift selective attention between separable acoustic dimensions. Moreover, cue-specific training (whether enhancing or inhibitory) appears to be more effective for the acquisition of second language speech contrasts.

3:51

5pSWb3. Investigating non-native category learning using a video-game-based training paradigm. Sung-joo Lim and Lori L. Holt (Dept. of Psychol., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213, sungjol@andrew.cmu.edu)

Adults have difficulty learning non-native speech categories, presenting an opportunity to study adult learning and plasticity with non-native speech categorization. Long-term training within laboratory-based response-feedback paradigms has produced modest non-native category learning in previous studies. The current study investigates the effectiveness of a video-game-based categorization training paradigm, found to be effective in learning novel nonspeech auditory categories [W. Holt, (2005)], to train native Japanese adults to categorize English /r/ and /l/. This approach emphasizes functional associations between sound categories and players' responses to video-game characters rather than overt phonetic categorization. Although categorization is not explicit in the game it is helpful to overall performance, providing a functional and perhaps more ecologically valid training signal than traditional feedback in standard laboratory training procedures. Japanese participants who played the game for about 2.5 h across 5 days with /r/ and /l/ sounds matched to game characters showed more nativelike perception of English /r/ and /l/ than control participants who played the game with nonspeech sounds. Listeners' categorization performance furthermore reflected regularities of the /r/-/l/ input distributions. In particular, variability across second-formant (F2) frequency may have encouraged listeners to shift perceptual cue weights toward more nativelike use of third formant (F3) frequency.

4:09

5pSWb4. Perception-production loop in native–non-native dialogs: Phonetic convergence. Natalie Lewandowski and Grzegorz Dogil (Inst. for Natural Lang. Processing, Univ. of Stuttgart, Azenbergstr. 12, 70174 Stuttgart, Germany, natalie.lewandowski@ims.uni-stuttgart.de)

The present study investigated the relation between the ability to adapt to the speech of a foreign language dialog partner and the individual variable of language talent. It was assumed that the actual degree to which one can converge to another speaker's pronunciation actually depends on his/her ability to pay attention to fine phonetic detail, to encode it, and to immediately reuse it in running speech—which was defined as phonetic talent. In order to analyze convergence in dialog, ten speakers of German (classified into two groups according to their proficiency and talent) performed a Diapix-task with a British and an American English conversational partner. Our main research goal can be described as follows: Does phonetic talent play a role in the perception-production loop and influence the process of accommodation towards a foreign dialog partner? The statistical analysis of the acoustic measurements revealed a significant effect for the factor talent across the American/British English conditions and also within the AE condition, indicating that the talented speakers have converged more. All results will be discussed assuming an underlying dynamic process, possibly involving a simultaneous interplay of convergence as well as divergence within a conversation.

4:27

5pSWb5. Second language influence on perception of first language phonotactics. Cynthia Kilpatrick (UCSD Linguist., 9500 Gilman Dr., #0108, La Jolla, CA 92093-0108)

The results of a speech perception experiment are reported, examining the responses of speakers of the same L1 (Spanish) with varying experience in the L2 (English). Three groups of speakers participated in a wordlikeness rating task in Spanish, in which nonce words were presented in sets that varied as to whether the final consonant of the word was phonotactically legal or illegal: (1) bilinguals who learned both English and Spanish before the age of 5, (2) L1 Spanish speakers that learned English after the age of 12, (3) monolingual speakers without communicative exposure to a second language. All groups differed significantly in the strength of their overall ratings, even though all speakers were technically native speakers of Spanish. In addition, for both the bilingual and second language speakers, ratings for legal vs illegal codas were significantly different. For monolingual speakers, on the other hand, ratings for legal and illegal codas were similar, but response times for these two conditions were significantly different. These results support the idea that native speaker judgements are affected by the extent of language experience that a speaker possesses, and the processing of phonotactic legality may not proceed similarly for monolingual, second language, and bilingual speakers.

Session 5pSWc

Speech Workshop: Keynote Address

Catherine T. Best, Chair
MARCS Auditory Labs., Univ. of Western Sydney, Penrith South DC, NSW 1797, Australia

Chair's Introduction—5:00

Invited Paper

5:05

5pSWc1. Automatic selective perception (ASP) of first-language (L1) and second-language (L2) speech: A working model. Winifred Strange (Speech Acoust. and Percept. Lab., CUNY-Graduate School, 365 Fifth Ave., New York, NY 10016-4309, strangepin@aol.com)

In this model, speech perception by adults is characterized as an active, information-seeking process whereby native listeners detect the most reliable acoustic parameters that specify phonetic segments and sequences, using highly over-learned, automatic selective perception routines. In laboratory perceptual tasks, differentiation of native (L1) phonetic contrasts is rapid and robust in suboptimal listening conditions even when the listeners focus on other levels of language structure, or indeed on another task. In contrast, late L2 learners must employ greater attentional resources in order to extract sufficient information to differentiate phonetic contrasts that do not occur in their native language. Phonetic and phonological modes of speech perception are described, which can be tapped in the laboratory by manipulations of stimulus complexity and task demands. These experimental manipulations reveal complex interactions between the linguistic experience of listeners and phonetic similarity relationships between L1 and L2 phonological inventories. Illustrative experimental evidence from studies of vowel perception using perceptual assimilation (cross-language identification), speeded discrimination, discrimination in speech babble, and brain indices of discrimination (MMN) will be presented to provide operational definitions of these concepts. Similarities and differences from other current theories of cross-language and L2 speech perception will be discussed. [Work supported by NIH, NSF.]

5:50—6:15 panel-discussion

Session 6aSWa

Speech Workshop: New Perspectives on Developmental Models

Ann R. Bradlow, Chair
Dept. of Linguistics, Northwestern Univ., Evanston, IL 60208

Chair's Introduction—9:00

Invited Papers

9:05

6aSWa1. Infant magnetoencephalography studies exploring neural links between sensory and motor representations for speech. Patricia K. Kuhl (Univ. of Washington, Inst. for Learning & Brain Sci., Mailstop 357920, Seattle, WA 98195-7920, pkkuhl@u.washington.edu)

The adult brain exhibits anatomical and functional specialization specific to speech, but we have little information regarding the infant brain. Recent adult neuroimaging studies show that speech processing is left-lateralized and that two regions of the brain, the superior temporal (ST, auditory area) and inferior parietal (IF, motor area), contribute to the brain's response to speech. To explore brain activation to speech and nonspeech in infants, we used magnetoencephalography to image the whole brain. Experiment 1 examined infants at three ages: newborn, 6 months, and 12 months. Infants showed significant ST activation to both speech and nonspeech stimuli. IF activation was not observed in newborns but in the two older groups of infants we observed synchronous IF and ST activation that was unique to speech. Experiment 2 investigated the nature of this perceptual-motor coupling for speech—specifically, we asked whether synchronous activation reflects the detection of speech signals per se, or whether it signals the recognition of experienced (native-language) speech. Native, non-native, and nonspeech analogs were used to test these alternatives. Our results elucidate the developmental time course of the sensory-motor connections for speech, and address longstanding theoretical issues in speech perception.

9:40

6aSWa2. Bilingual beginnings as a lens for theory development. Janet F. Werker, Suzanne Curtin, and Krista Byers-Heinlein (Dept. of Psych., Univ. of British Columbia, 1236 West Mall, Vancouver, BC V6T 1Z4, Canada, kbyers@psych.ubc.ca)

Phonetic categories become language specific across the first months of life. However, at the onset of word learning there are tasks in which infants fail to utilize native language phonetic categories to drive word learning. In 2005, we advanced a framework to account for why infants can detect and use phonetic detail in some tasks but not in others (Werker and Curtin, 2005; see also Curtin and Werker, 2007). In this framework, PRIMIR (processing rich information from multidimensional, interactive representation), we argue that by their first birthday, infants have established language-specific phonetic category representations, but also encode and represent both subphonetic and indexical details of speech. Initial biases, developmental level, and task demands influence the level of detail infants use in any particular experimental situation. On some occasions phonetic categories are accessed, but in other tasks they are not given priority. Recently, we have begun studying infants who are exposed to two native languages from birth (Werker and Byers-Heinlein, 2008). In the current paper, we will review recent work on speech perception and word learning in bilingual-learning infants. This will be followed by a discussion of how this research has lead to advances in, and changes to, PRIMIR.

SATURDAY MORNING, 23 MAY 2009 SKY BRIDGE AUDITORIUM, 10:30 A.M. TO 12:15 P.M.

Session 6aSWb

Speech Workshop: Early Bilingual Development

Susan G. Guion, Chair
Dept. of Linguistics, Univ. of Oregon, Eugene, OR 97403-1290

Invited Papers

10:30

6aSWb1. Phonetic variability in bilinguals' acquisition of native-vowel category contrasts. Laura Bosch and Marta Ramon-Casas (Dept. of Basic Psych., Univ. of Barcelona, 171 P. Vall d'Hebron, Barcelona 08035, Spain)

Previous research has shown that bilingual infants and toddlers follow a slightly different developmental pattern in building and stabilizing some native vowel contrastive categories, compared to monolingual populations. Once in the lexical stage, bilingual toddlers do not always show the expected mispronunciation effect when presented with familiar words in which a vowel change has been introduced. To better understand bilinguals' protracted process in setting the boundaries for certain, perceptually close, native vowel categories, a phonetic analysis of the input characteristics was undertaken. Speech samples containing target words with Catalan mid-front vowel sounds were recorded from two groups of Catalan-speaking mothers from monolingual and bilingual environments, differing in the use of their L2 (Spanish). First and second formant values of the target vowels were obtained and compared between groups. The vowels were always correctly and contrastively produced, but bilinguals' formant values differed significantly from monolinguals': categories were closer in space and greater variability in production was found. Young bilinguals' sensitivity to phonetic variability in the words they hear may constrain the development of neat representations of contrastive vowel categories in the early lexicon. [Work supported by CONSOLIDER 2010 Program (CSD2007-012).]

11:05

6aSWb2. How are speech sound categories acquired in bilingual first language acquisition? Megha Sundara and Adrienne Scutellaro (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095-1543, megha.sundara@humnet.ucla.edu)

For bilingual infants, phonetic categories of the two languages develop interdependently. This is supported by the differences in time course and trajectory of development of phonetic perception between Spanish-Catalan bilingual and monolingual infants (Bosch and Sebastián-Gallés, 2003a, 2003b, 2005; Sebastián-Gallés and Bosch, in press). Bosch and Sebastián-Gallés argue that for similar phones, at least initially, infants track statistical regularities across the two languages. This results in a temporary inability of bilingual infants to discriminate acoustically similar phonemes. Research on French-English bilingual infants shows a different pattern. There is evidence that French-English bilingual infants may not differ from their monolingual peers in discrimination of phonetic contrasts (Burns *et al.*, 2007; Sundara *et al.*, 2008). In this talk, data will be presented from Spanish-English bilingual infants to evaluate how development of phonetic perception is affected by (a) frequency of phones and (b) the rhythmic (dis)similarity of the two languages to which bilinguals are exposed.

11:40

6aSWb3. The development of English vowel perception in monolingual and bilingual infants: Neurophysiological correlates. Valerie L. Shafer (Program in Speech-Lang.-Hearing Sci., The Graduate Ctr., CUNY, 365 Fifth Ave. NY, NY 10016)

This research examines the maturation of I-E vowel discrimination, in terms of neurophysiological responses, in infants exposed monolingually to English or bilingually to English and Spanish. The vowels [I] as in "pit" and [E] as in "pet" contrast meaning in English but not in Spanish. Bilingual exposure to English and Spanish alters the distribution of phonetic information heard by an infant and may alter the timecourse of vowel category development in the two languages [e.g., L. Bosch and N. Sebastian-Galles (2003)]. The goal of the current study is to examine whether there is evidence from brain responses indicating differences in the development of this English vowel category as a function of amount of English versus Spanish in the input. Preliminary results reveal that both monolin-

gually and bilingually exposed infants and children show evidence of good discrimination of the English vowels, but that the groups with bilingual exposure more often show increased contribution of the right hemisphere to discrimination compared to the groups receiving monolingual exposure. It is not currently clear why these hemispheric differences are found, but they may be related to differences in how attention is allocated in processing the vowels. [This research was supported by NIH HD46193.]

SATURDAY AFTERNOON, 23 MAY 2009 SKY BRIDGE TERRACE, 1:30 TO 3:00 P.M.

Session 6pSWa

Speech Workshop: Cross-Language Speech Perception and Linguistic Experience: Poster Session C

Suzanne L. Curtin, Chair
Dept. of Psychology, Univ. of Calgary, Calgary, AB T2N 1N4, Canada

Contributed Papers

All posters will be on display from 1:30 p.m. to 3:00 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 2:15 p.m. and contributors of even-numbered papers will be at their posters from 2:15 p.m. to 3:00 p.m.

6pSWa1. English vowel perception by native speakers of European Portuguese and Mandarin. Andreia S. Rauber (Ctr. for Humanistic Studies, Univ. of Minho, Campus de Gualtar, Braga, Portugal 4715-057, arauber @ilch.uminho.pt)

This study investigated the perception of eight American English stressed syllabic monophthongs (/i, I, E, æ, V, O, U, u/) by two groups of nonnative listeners: Portuguese and Chinese. Each group was formed by 15 English as a foreign language (EFL) learners, with an upper-intermediate level of English proficiency, all undergraduate students at a Portuguese university. An identification test was designed to investigate how the participants would perceive the English vowels, which are not present in their first language (L1) inventory as stressed monophthongs (/I, æ, V, U/ in the case of the Portuguese speakers, and all vowels but /i, u/ in the case of the Mandarin speakers). The large number of target vowels played as stimuli allowed a cross-language analysis of vowel distribution in the acoustic space: for the two groups of EFL listeners, it was more difficult to perceive the English vowels located in a high density area of the English vowel space than those in less crowded areas. For the two groups of speakers, the vowels /æ/ and /U/ had the greatest misidentification rates.

6pSWa2. Does the native language use of duration affect the perception of non-native length contrasts? Yukari Hirata (Dept. of East Asian Lang. and Lit., Colgate Univ., 13 Oak Dr., Hamilton, NY 13346, yhirata@colgate.edu) and Motoko Ueyama (Univ. di Bologna, 47100 Forli, Italy)

This study compared native Italian (NI) and American English (NE) speakers' abilities to perceive Japanese phonemic length contrasts. Japanese has both vowel and consonant length contrasts, Italian has only consonant length contrast, and English has neither. The study examined to which extent the differential use of duration in their native languages affects their abilities to perceive length contrasts in an unfamiliar language. Twenty-two monolingual NI speakers perceived Japanese words in isolation and in sentences and were asked to identify the number of moras in target words, e.g., /o.do.t.ta/ as four moras. Their test scores were compared to those of 18 NE participants [Hirata (2004)] in an analysis of variance. Group (NE, NI) was a between-subjects factor, and context (isolation, sentences) and word type (long vowels, geminates, combination, and short segments) were within-subjects factors. While there was no main effect of group (NI: 44.6%; NE: 39.5%), there was a significant three-way interaction. The NI group scored significantly higher than the NE group on words with geminates spoken in sentences (43.3 vs 31.1%). Results are discussed as to whether the perception of non-native length contrasts is driven by language-specific ability [McAllister *et al.* (2002)] or by more general auditory ability [Bohn (1995)].

6pSWa3. Linguistic experience in tone perception. James Kirby (Phonology Lab., Dept. of Linguist., 1010 E 59th St., Chicago, IL 60637, jkirby@uchicago.edu)

This paper examines the effect of regional dialect on tone perception. Thirty speakers of Northern and Southern Vietnamese performed an AX discrimination task using natural speech tokens of Northern (NVN) speech. NVN distinguishes six tones, three of which are produced with creaky voice, while Southern Vietnamese (SVN) distinguishes only five tones, none of which are canonically produced with creaky voice; however, both groups of listeners have shown some sensitivity to voice quality as well as F0 cues (Brunelle, 2008). While the results of hierarchical cluster analysis show both Northern and Southern listeners broadly group tones by F0 onset, multidimensional scaling shows the effects of dialect-specific perception: despite some familiarity with NVN speech, SVN listeners did not adjust their perceptual cue space when listening to NVN talkers. As a result, tones with similar F0 profiles but different voice qualities were more likely to be confused by SVN listeners. This is consistent with studies showing that perceptual processes are organized in a language-specific fashion, with the acoustic cue space weighted by phonological relevance to L1 perception (Werker and Tees, 1999; Strange, 2002). It is argued that language-specific prosodic as well as segmental experience affects speech processing at the prelexical level.

6pSWa4. Cross-linguistic interpretation of duration. Ellen Broselow, Jiwon Hwang (Dept. of Linguist., Stony Brook Univ., Stony Brook, NY 11794-4376, ellen.broselow@stonybrook.edu), and Nancy Squires (Stony Brook Univ., Stony Brook, NY 11794)

In Korean, intervocalic [l] is realized as tap ([tal]/[tar-i] 'moon /moon+nom'). In English loanwords, however, intervocalic /l/ is generally adapted as a geminate lateral ([sollo] 'solo' but [sara] 'Sarah'). We present evidence from event-related potentials supporting an analysis in which Korean listeners perceive intervocalic single [l] (illegal in Korean) as geminate [ll], reinterpreting the English [r-l] contrast in terms of the Korean [r-ll] contrast ([dari] 'bridge', [dalli] 'differently'). Korean and English participants heard two sets of oddball paradigms, [ele-elle] and [ene-enne]. In both cases, the acoustic difference is the same, 48 versus 98 msec. However, the nasal pair represents a cross-category contrast in Korean ([kanan] 'poverty', [kannan] 'newborn') while the lateral pair represents a noncontrastive difference. Consistent with studies showing a stronger mismatch negativity (MMN) to cross-category changes than to within-category changes, Korean listeners displayed a significantly larger MMN for the nasal pair than for the lateral pair. In contrast, English listeners (for whom the pairs do not differ in categorical status) did not show significantly different responses to the durational changes in nasals versus laterals. Korean listeners' relatively weak

response to durational differences in laterals suggests that their adaptation pattern reflects inaccurate mapping of the acoustic signal to phonological representations. [Work supported by NSF BCS-07460227.]

6pSWa5. Informational masking in first- and second-language speech recognition. Kristin J. Van Engen (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, k-van@northwestern.edu)

Human speech recognition in noisy conditions is modulated by cognitive factors such as language background. For example, noise is more detrimental to non-native listeners than native listeners (e.g., van Wijngaarden *et al.*, 2002), and when noise is a speech signal, native-language noise is more detrimental than foreign-language noise for listeners attending to native-language speech targets (e.g., Van Engen and Bradlow, 2007). It is not clear, however, whether this increased interference is primarily due to the native status of the noise or to the greater similarity between target and noise. To address this issue, English speech recognition in the presence of English and Mandarin babble was assessed for monolingual English listeners and L2 English listeners whose L1 is Mandarin. Results showed that intelligibility for both groups was lower in English versus Mandarin babble; that is, L2 listeners experienced more difficulty in same-language noise versus native-language noise. However, monolingual English listeners showed a greater release from masking in Mandarin noise than did L1 Mandarin listeners. The informational masking imposed on speech targets by interfering speech noise, therefore, is dependent both on the linguistic and/or acoustic similarity between the noise and the target and on the listeners' experience with the languages involved.

6pSWa6. Discrimination of English, French, and Spanish liquids by Japanese listeners. Tomohiko Ooigawa (Graduate School of Foreign Studies, Phonet. Lab., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, Japan, ooigawaferchichi@gmail.com)

The present study examines the discrimination of English, French, and Spanish intervocalic liquids' contrasts (English /l/-/r/, French /l/-/r/, and Spanish /l/-/r/, /l/-/rr/, /l/-/rr/) by native speakers of Japanese. The results show that the Japanese listeners discriminated the contrasts of French liquids and Spanish trill versus lateral/tap at the rate of more than 95%. On the other hand, they poorly discriminated the contrasts of English liquids and Spanish tap versus lateral. This study discusses whether the current models on second language speech perception account for the phenomena. It has been said that native speakers of Japanese have difficulty in distinguishing the contrasts of liquids both productively and perceptually. A lot of empirical studies on the perception of English liquids by Japanese listeners have been carried out so far. However, there are few empirical studies on the perception of other languages' liquids by Japanese listeners. This is the first study which compares the discrimination of the liquids of the three languages by native speakers of Japanese. In the experiment, utterances of /peCV/ were recorded from the native speakers of the three languages. AXB task was used for the perception experiment.

6pSWa7. Discrimination of four English vowel contrasts by Catalan learners varying in language experience. Lucrecia Rallo Fabra (Univ. of the Balearic Islands, ctra. Valldemossa, km. 7,5, 07122 Palma de Mallorca, Spain)

PAM-L2 (Best and Tyler, 2007) hypothesizes that discrimination of L2 sounds can be predicted from the perceptual relatedness of L2 categories to L1 categories. Catalan EFL learners encounter serious difficulties to discriminate some English vowel contrasts, partly because Catalan has a smaller vowel inventory than American English. This study presents data from a perceptual discrimination test in which three groups of EFL learners varying in experience with English were asked to discriminate a series of potentially difficult English vowel pairs. The discrimination task consisted in picking the odd item out of three stimuli which were heard one after the other at 1.3-s intervals. The target vowel stimuli occurred in CVC syllables produced by six native English talkers. Four vowel contrasts were tested: /*opena-/invv*/, /*eh-/aelig*/, /i-/*smallI*/, and /u-/*smallU*/. As predicted by PAM-L2, "category-goodness" contrasts /i-/*smallI*/ and /u-/*smallU*/ were fairly well discriminated, the "single-category" contrast /*opena-/invv*/ was poorly discriminated. The "uncategorized" contrast /*eh-/aelig*/ failed to meet PAM-L2 predictions, learners failed to distinguish these two vowels although the contrast was predicted as relatively easy to discriminate. Language experience did not have a significant effect on ease of discrimination. These results suggest that PAM-L2 assimilation patterns can also be extended to EFL learning.

6pSWa8. Language preference in monolingual and bilingual infants. Linda Polka, Ayasha Valji (School of Commun. Sci. and Disord., 1266 Pine Ave. W., McGill Univ., Montreal, QC H3G 1A8, Canada, linda.polka@mcgill.ca), and Karen Mattock (Lancaster Univ., Fylde College, Bailrigg, Lancaster LA1 4YW, UK)

Previous research shows that infants in single-language families have some basic language discrimination abilities at birth which improve rapidly over the first 6 months of life, and that attention to the rhythmic properties of language supports these skills. Babies in monolingual families also prefer listening to their native language over an unfamiliar language when presented samples produced by one bilingual or two monolingual talkers. In this study we investigate the emergence of language-specific speech processing in bilingual infants by comparing language preference patterns in monolingual English, monolingual French, and bilingual English-French infants using a three-way language preference test. Listening times were measured to passages of adult-directed speech from three rhythmically different languages (English, French, Japanese; three talkers per language). Ten-month-olds in the monolingual groups listened equally to all three languages. However, 10-month-old bilinguals showed a significant preference for each native language over Japanese; listening times to English and French were not different. Individual bilingual 10-month-olds preferred the more prevalent native language in their input. These findings indicate bilingual infants listen more selectively when they encounter different languages. The implications of these findings for understanding speech processing in early bilingual acquisition will be discussed. [Work supported by SSHRC.]

6pSWa9. The perception of Georgian ejective stops by native English speakers. Christopher S. Doty and Susan G. Guion (Dept. of Linguist., 1290 Univ. of Oregon, Eugene, OR 97403-1290, cdoty@uoregon.edu)

The present study examined the perception of Georgian voiced, voiceless aspirated, and ejective stops by native speakers of English. It was motivated by the observation that languages with three stop-manner series often borrow voiceless aspirated stops as ejectives in a systematic manner. This is likely due to the perceptual similarity between aspirated stops in source language and the ejective stops in the borrowing language. To test the hypothesized perceptual similarity between aspirated and ejective stops, five English speakers were asked to listen to stops which varied in manner of articulation (voiced, voiceless aspirated, ejective) that were produced by a speaker of Georgian. Participants heard three productions at the same place of articulation, one of which differed in manner, and were asked to identify the the oddball stop. The results indicated that the contrast between voiceless aspirated and ejective stops was correctly identified less often than the other two contrasts (voiced versus aspirated, voiced versus ejective) [$p<0$ .05]. No significant difference was found for the other contrasts. These data are supported by the results from a second task, in which the same participants performed a forced-choice identification and goodness rating task with the same productions.

6pSWa10. Perception-production relationship in French vowel learning in adulthood. Erika S. Levy (Speech-Lang. Pathol., Biobehavioral Sci., Teachers College, Columbia Univ., 525 W. 120th St., Box 180, New York, NY 10027, elevy@tc.columbia.edu) and Franzo Law, II (City Univ. of New York-Grad. Ctr., New York, NY 10016-4309)

The present study examined the extent to which perceptual performance by American English (AE) individuals predicted their accuracy in producing second-language (L2) Parisian French (PF) vowels. Three groups of AE participants (no, moderate, and extensive French-language experience) participated in two perceptual tasks (categorial discrimination and perceptual assimilation), and a production (repetition) task involving PF /y-œ-i-a-u/ in bilabial /rabVpa/ and alveolar /radVta/ contexts within a phrase. Results from perception tasks correctly predicted overall production difficulties and effects of language experience and consonantal context in L2 production. Paralleling their perceptual patterns, front rounded vowel productions by AE participants were mislabeled more often as back rounded vowels than as front vowels by native-French speakers. PF /œ/ was produced more accurately with greater L2 experience. Production accuracy of /y/ was also

greater with extensive experience, a finding not expected based on minimal language-experience effects on /y/ perception. Productions of PF /u/ were identified approximately twice as accurately in bilabial context as in alveolar, in a pattern consistent with discrimination performance. Thus, PF /u/ may be considered a "similar" vowel in alveolar context, but not in bilabial, suggesting an allophonic level of representation in L2 learning, both in perception and production.

6pSWa11. Time-course of perception of Mandarin Chinese tones. Chiung-Yun Chang and Robert Allen Fox (Speech Percept. and Acoust. Labs., Speech and Hearing Sci. The Ohio State Univ., Columbus OH 43210-1002, chang.553@osu.edu)

The current study examines the effects of dialect variation on the time-course of lexical tone identification by native and non-native listeners of Mandarin Chinese. Listeners were asked to identify the tone of an isolated Chinese word presented without a precursor phrase. All stimuli were exemplars of V, CV, or CVC combinations that represented real words in Mandarin when produced with any of the four possible lexical tones. Words were produced by multiple talkers with a wide range of F0 values. The dialect of the talkers was also systematically varied. Half were native speakers of Beijing Mandarin, and the other half were native speakers of Taiwanese Mandarin. Preliminary acoustic analysis showed significant differences in tonal realization of the four lexical tones, especially Tone 3. There were three groups of listeners: native speakers of Beijing Mandarin and Taiwanese Mandarin and native English speakers who were learning Beijing Mandarin. A gating task was utilized to determine the temporal location in the token when the lexcial tone was correctly identified and to examine the nature (and time course) of tone confusions. Results will be discussed in terms of the effects of dialect and F0 variation of the stimuli and the native language of the listener.

6pSWa12. Perception or production? Training effects on cross-language phonological awareness tasks in Mandarin-speaking children. Pi-Yu Chiang and Susan Rvachew (School of Commun. Sci. and Disord., McGill Univ., 1266 Pine Ave. West, Montreal QC H3G1A8, Canada, piyu.chiang@mail.mcgill.ca)

This study investigated training effects of perception-based and production-based English activities on the acquisition of phonological awareness of English sound structures by 58 Mandarin-speaking kindergarten-aged children in Taiwan. Children were randomly assigned to one of two experimental conditions or a control condition. Experimental groups participated in a learning session of four English words: band, nest, brick, and stool, with a perceptual or articulation focus on clusters, which do not exist in Mandarin. Outcome measures examined subjects' ability to match words on the basis of shared onset or coda, and to elicit common units in both Mandarin and English before and after training. Significant gain in English onset common unit test scores was observed for the production training group, relative to the control group, especially in trials containing the trained clusters: br- and st-. Perception group however did not show an improvement over time across all measures. Between-group differences were not observed for English codas, Chinese onsets, or Chinese codas. Findings revealed that clearly articulating non-native words with unfamiliar syllable structures could assist children to isolate onsets in explicit phonological awareness tasks, independent of exposure to alphabetic orthography, generally thought to be critical in explicit phonological awareness (e.g., Gombert, 1992).

6pSWa13. Learn to Listen (L2L): Perception training system for learners of English as a second language. Diane Kewley-Port (Commun. Disord. Technol., Inc. and Indiana Univ., Bloomington, IN), Kanae Nishi (Commun. Disord. Technol., Inc. and Boys Town Natl. Res. Hosp., Omaha, NE), Hanyong Park, James D. Miller, and Charles S. Watson (Commun. Disord. Technol., Inc. and Indiana Univ., Bloomington, IN)

Computer software (L2L) is being developed for comprehensive perception training of English by second language learners. Our goal is to facilitate generalization of post-training improvement of phoneme perception to the perception of running speech. Three studies are reported for two groups of adult listeners, one Korean and the other Spanish. In study 1, large sets of confusable phonemes were identified from an assessment task for each group. Then training sets for consonants in CV nonsense syllables and vowels in familiar real words were selected and recordings from multiple talkers were obtained. Materials for the word-in-sentence (WIS) task were developed with a single low-context carrier phrase which contained three words from the vowel training. In study 2 new Korean and Spanish listeners were trained using a protocol that included a pre-test, eight hours of training, a post-test, and one hour using the WIS task. Training results showed: (1) both Korean and Spanish listeners improved from pre- to post-tests for vowels and consonants; (2) in vowel identification, Koreans outperformed Spanish by 34% (3) vowel identification in the WIS task appeared easier than in word-only task, a result confirmed by a study using the WIS task without training (study 3). [Work supported by NIH.]

6pSWa14. Effects of linguistic and musical experience on non-native perception of Thai vowel duration. Angela Cooper and Yue Wang (Dept. of Linguist., Simon Fraser Univ., 8888 Univ. Dr., Burnaby, B.C. V5A 1S6 Canada, akcooper@sfu.ca)

Previous research has suggested a relationship between musical experience and L2 proficiency. The present study investigated the influence of musical experience on non-native perception of speaking-rate varied Thai phonemic vowel length distinctions. Given that musicians are trained to discern temporal distinctions in music, we hypothesized that their musical experience would enhance their ability to perceive non-native vowel length distinctions as well as their sensitivity to temporal changes as a function of speaking rate. Naive native English listeners of Thai, with and without musical training, as well as native Thai listeners, were presented with minimal pairs of monosyllabic Thai words differing in vowel length at three speaking rates in an identification task and a discrimination task. For identification, participants were asked to identify whether a word contained a long or short vowel. For discrimination, participants heard three successive words and were asked to indicate whether the second word had the same vowel length as the first or last word. The results show significant group differences in identification and discrimination accuracy within and across speaking rates, suggesting that listeners' perception of phonetic categorical versus temporal acoustic variations differs as a function of linguistic and musical experience.

6pSWa15. Increased second-language vocabulary size is associated with more consistent second-language vowel categorization and better discrimination. Rikke L. Bundgaard-Nielsen (MARCS Auditory Lab., Univ. of Western Sydney, Bankstown Campus, Bldg. 1, Locked Bag 1797, Penrith South DC, NSW 1797, Australia, r.bundgaardnielsen@uws.edu.au), Catherine T. Best, and Michael D. Tyler (Univ. of Western Sydney, NSW 1797, Australia)

Improvement in second-language (L2) perception has been posited to asymptote early in L2 learning when the L2 vocabulary is still small, whereas a large L2 vocabulary curtails perceptual learning [PAM-L2: Best and Tyler (2007)]. We extend this proposition by suggesting that early L2 lexical development facilitates establishment of phonological categories in a manner analogous to toddlers' L1 acquisition before versus after the vocabulary explosion. According to this revised view, L2 speech should be assimilated more consistently to phonological categories, and cross-boundary contrasts discriminated more accurately, by learners with large versus small L2 vocabularies. We applied a novel whole-system approach to evaluate perception of L2 vowels with respect to the L1 vowel system. Japanese learners of Australian English with under 12 weeks of L2 immersion completed labeling and goodness ratings on all Australian vowels, using all mono- and bi-moraic Japanese vowels and vowel combinations. They also completed discrimination tasks on four L2 vowel contrasts, representing a range of PAM-L2 contrast types, and a L2 vocabulary size assessment. Learners with larger vocabularies had more consistent vowel assimilation and more accurate cross-boundary discrimination than those with smaller vocabularies, supporting the proposition that lexical development supports L2 phonological acquisition. [Work supported by NIH and MARCS/UWS.]

6pSWa16. The perception of Norwegian word tones by second language speakers. Wim A. van Dommelen and Olaf Husby (Dept. of Lang. and Comm. Studies, NTNU, N-7491 Trondheim, Norway, wim.van.dommelen@hf.ntnu.no)

The present study is concerned with the perception of Norwegian word tones by second language speakers. The phonological system of Norwegian involves two different tones that are used to distinguish a relatively large

number of word pairs. Earlier studies have shown that native speakers of Norwegian are able to identify the word tones almost error-free. The aim of the present study was to investigate word tone perception by speakers of a tonal language (Mandarin Chinese) and a non-tonal language (German). These two groups as well as a control group of native listeners identified manipulated stimuli with tonal contours varying between tone 1 and tone 2. It appeared that the L2 users had less sharp transitions than the natives, the Chinese subjects performing somewhat better than the Germans. In addition, native speakers categorized tokens of tone 1 and tone 2 words produced by the L2 speakers and indicated on a five-point scale how sure they were in their judgment. In an acoustical analysis the fundamental frequency contours of the L2 test words were analyzed and subsequently used for comparison with native categorization results. The implications of the found correlations between production and perception are discussed.

6pSWa17. Prominence perception: Conflicting cues and linguistically encoded bias. Miran Kim (Dept. of Linguist., State Univ. of New York at Stony Brook, S201 SBS Bldg., Stony Brook, NY 11794, mrkim@ic.sunysb.edu)

Beyond the commonly known fact that prominent syllables are greater in amplitude, higher in F0, and longer in duration, this study examines cross linguistic patterns in the perception of prominence. Particularly, identification of stress placement is tested in a situation where F0 and duration happen to be in conflict (e.g., H+L* bitonal pitch accent). Three languages are selected considering their prosodic characteristics such as phonological role of prominence and rhythm (stress-timed/syllable-timed): English, Spanish, and Korean. A new-language-learning setting is devised to collect production data based on which the patterns of phonetic realization of stress are directly compared among the language groups. CVCVCV pseudowords, where C=m, n, k, s and V=a, (e, i), are used for both production and perception experiments. F0 and duration are manipulated in order to create conflicting cues that are suitable for the purpose of the perception experiment. Stressed syllables are consistently realized with higher F0 together with longer duration across the groups, though to different extents. The perceptual sensitivity to the durational manipulation is found to vary among the groups. Linguistic implications are discussed referring to the phonological role(s) of F0 and duration in each language as well as to the relationship between production and perception.

6pSWa18. Perceptual attunement in infants at risk of reading disabilities. Christine Kitamura and Anna Herald (MARCS Auditory Lab., Bldg. 5, Bankstown Campus, Univ. of Western Sydney, Locked bag 1797, Penrith South DC, NSW, 1797, c.kitamura@uws, edu.au)

Impaired phonological processing has been found to have a reciprocal casual association with the reading ability of people with reading difficulties. Further, there is growing evidence that problems in phonological processing are present at birth as research shows that infants with a family history of these disorders have atypical neural electrical responses (ERPs) to speech stimuli. However, the nature of the phonological deficit is not clear. This study examined the phonological development of at-risk 9- to 12-month-old infants (n=18) and a control group (n=18). Their discrimination performance was evaluated using the habituation-dishabituation task to test their ability to discriminate the confusable native contrast, /fa-tha/ and the non-native contrast, /k'i-q'i/ from Werker and Tees (1984). In line with core phonological deficit hypothesis and hereditary accounts of reading difficulties, infants' attunement to their native language was significantly correlated with parental phonological ability measured using Pseudoword Decoding Test of the WIAT-II. The at-risk infants easily discriminated both native and non-native contrasts, whereas the control infants discriminated the native contrast but could not discriminate the non-native one. These results indicate that at-risk infants do not attune to their native language at the same rate as their peers.

6pSWa19. Effects of visual cues and phonetic contexts in perception of non-native contrasts by Cantonese learners of english. Bin Li (Dept. of Chinese, Translation & Linguist., City Univ. of Hong Kong, Hong Kong SAR, China)

Visual articulatory information, in addition to audio features, is integrated in L1 and L2 speech perception automatically and unconsciously [McGurk and MacDonald (1976); Rosenblum *et al.* (1997); Hardison (2003); Hazan *et al.* (2005)]. Previous literature on consonant perception has reported that visual aid is most significant with places of articulation, and that with manners, except with /r/ and /l/, has received little attention due to lack of research significance. Contrary to the commonly held idea, however, that the articulation of [l] and [n] is visually similar, our examination on video recording of three native English speakers' production of words contrasting [l] and [n] syllable initially suggests visual differences in advancement of tongue tips. This study investigates effects of such visual information in the identification of the non-native contrast by Cantonese learners of English. Results show that the relatively more distinct visual information can only help Cantonese speakers better perceive the two sounds in certain contexts, and in others cause more confusion.

6pSWa20. Do words in the native language influence second-language speech perception? Evidence from Japanese listeners' perception of English words that exist as Japanese loanwords. Keiichi Tajima (Dept. of Psych., Hosei Univ., 2-17-1 Fujimi, Chiyoda-ku, Tokyo 102-8160, Japan, tajima@hosei.ac.jp)

It is well known that native-language (L1) sound structure influences adult learners' perception of second-language (L2) speech. However, it is not as clear whether L2 speech perception can also be influenced by L1 mental lexicon. Specifically, L2 words may be inaccurately perceived if familiar loanword counterparts exist in L1 and if these loanwords have different sound patterns from the original words. To examine this, the present study took advantage of Japanese loanwords from English (e.g., Japanese /sutoresu/ from English "stress"), which typically undergo syllable structure changes, and investigated whether Japanese listeners can accurately perceive the syllable structure of English words even when loanwords exist in Japanese. Thirty native Japanese listeners were asked to count syllables in 189 English words spoken by a native English talker, and rate their subjective familiarity with the English words and their loanword counterparts transcribed in Japanese katakana orthography. Results indicated that syllable-counting accuracy (65% on average) was not influenced by familiarity with the English words nor with the Japanese loanwords. Results suggest that even though loanwords resemble the original words in their sound and meaning, knowledge of such words in L1 does not necessarily influence perceptual processing of L2 words. [Work supported by MEXT and JSPS.]

6pSWa21. Predicting second language (L2) identification rates from first language (L1) mapping data: Similarity patterns for English and Korean obstruents in pre- and poststressed intervocalic, and postvocalic positions. Hanyong Park (Speech Res. Lab., Indiana Univ., 1101 E. 10th St., Bloomington, IN 47405, hanypark@indiana.edu) and Kenneth J. de Jong (Dept. of Linguist., Indiana Univ., Bloomington, IN 40405)

Park and de Jong [J.Phon, **36**, 704–723 (2008)] found that listeners' identification rates of second language (L2) categories can be predicted from mapping data, provided the L2 category has a high degree of subjective correspondence to native language (L1) categories. The current study examines whether identification rates can be predicted for consonants in different prosodic locations, and hence, whether the reliance on L1 categories is the same, regardless of prosodic position. Forty native Koreans identified /p b t d f v θo/ in pre- and poststressed intervocalic, and postvocalic positions, in nonsense English words, using both Korean and English labels, and also gave gradient evaluations of their identifications. The mapping data were used to predict confusion rates for the L2 identification data, finding results for prestress intervocalic consonants similar to the previous study for word-dinitial consonants; stop identification rates were well-predicted, but fricatives are systematically much better than predicted. In poststress position, however, fricative identification rates were very well predicted, suggesting that fricatives were treated as "similar" to L1 categories. These results indicate that the degree of assimilation of an L2 to an L1 category is heavily dependent on prosodic location. [Work supported by NSF BCS-04406540.]

6pSWa22. Perception of foreign accent in Japanese and Brazilian Portuguese: Prosodic and segmental factors. Masahiko Komatsu (School of Psychol. Sci., Health Sci. Univ. of Hokkaido, Ainosato 2-5, Sapporo, 002-8072 Japan, koma2@hoku-iryo-u.ac.jp)

Japanese sentences read by Japanese-descent Brazilians (L2) and native Japanese speakers (L1) were rated by native Japanese speakers, and Portuguese sentences read by native Japanese speakers learning Portuguese (L2)

and Japanese-descent Brazilians (L1) were rated by Brazilians. The raters evaluated three types of samples: (a) sounds without spectral properties (representing prosodic features); (b) *F*0-flattened sounds (representing segmental features); and (c) unmodified sounds. In both languages, the perceptual scores of L1 and L2 samples were separated most clearly in (c), followed by (b) and (a). Both the scores of (a) and (b) showed strong correlations with (c) in L2. These suggest that both prosodic and segmental features play a role in the detection of foreign accent, but the latter has a greater effect. *F*0 characteristics varied between speaker groups rather than languages. The speaking rate in L2 was correlated with the perception scores. The effect of segmental features was greater in L2 Portuguese than in L2 Japanese, i.e., greater in less fluent speech. Foreign accents in both languages share common properties although having some differences. [This work is the revised version of Komatsu and Kimoto (2008) and Komatsu (in press), and partially supported by KAKENHI (20242010) and HSUH Kotaisa grant.]

6pSWa23. Acquisition of new phonetic categories by bilingual and monolingual groups: Role of metalinguistic awareness and feature generalization. Divya V. Gogoi (Linguist. Program, Univ. of Florida, P.O. Box 115454, Gainesville, FL 32611) and James D. Harnsberger (Dept. of Commun. Sci. Disord., Univ. of Florida, Gainesville, FL 32611)

The present study examines the role of feature generalization and metalinguistic awareness in the perception and learning of novel non-native speech contrasts by 40 bilingual speakers (Bengali–English and Spanish–English) and 20 monolingual speakers of American English. Bilinguals, unlike monolinguals, employ certain cognitive and linguistic skills during lexical processing and word learning attributed as metalinguistic awareness. This awareness in turn, can be applied in mastering unfamiliar phonetic features in a third language. A second issue concerns the role that native phonetic features may play in the development of new phonetic categories involving same phonetic features (feature generalization hypothesis). To explore both factors, a high variability perception training paradigm was used to examine the acquisition of novel non-native speech contrasts (containing dental/alveolar-retroflex place distinction in various manners) from a target language, Malayalam. A consonant identification procedure was used throughout the training and testing phases of the experiment to examine post-test and generalization performance across groups. Productions of the stimuli were also recorded at the pretest and posttest phases. In addition, perceptual assimilation patterns were analyzed from consonant assimilation and discrimination scores before and after training.

6pSWa24. Dutch listeners' perception of Korean stop triplets. Mirjam Broersma (Radboud Univ. Nijmegen, Donders Inst. for Brain, Cognition and Behaviour, Ctr. for Cognition, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands, mirjam@mirjambroersma.nl)

This study investigates Dutch listeners' perception of Korean stop triplets. Whereas Dutch distinguishes prevoiced and voiceless unaspirated stops, Korean distinguishes fortis, lenis, and aspirated stops. Here, perception of fortis, lenis, and aspirated bilabial (/pp/-/p/-/ph/), alveolar (/tt/-/t/-/th/), and velar (/kk/-/k/-/kh/) stops is investigated. In Dutch, VOT is the most important perceptual cue for initial stop voicing. Korean fortis stops fall within the VOT range of Dutch voiceless stops; VOTs of lenis and especially aspirated stops are longer than Dutch VOTs. Therefore, Dutch listeners were expected to distinguish fortis stops more accurately from the other two than lenis and aspirated stops from one another. In a phonetic categorization experiment, Dutch listeners categorized Korean stops in naturally recorded CVs as fortis/lenis, lenis/aspirated, or fortis/aspirated. Indeed, for all places of articulation, fortis stops were distinguished relatively accurately from the other two stops. The most difficult distinction was between lenis and aspirated stops (both outside the Dutch VOT range), the easiest distinction between fortis and aspirated stops (the former within the Dutch VOT range, the latter the most remote from Dutch VOTs). Thus, although all Korean VOT category boundaries are outside the Dutch VOT range, the distance from Dutch VOT values affected Dutch listeners' categorization.

6pSWa25. Effects of native language and amount of experience on crosslinguistic perception. Juli Cebrian (Dept. Filologia Anglesa, Fac. Filosofia i Lletres-Edifici B, Univ. Autonoma de Barcelona, Bellaterra 08193, Spain, juli.cebrian@uab.es)

Models of second language (L2) speech learning draw on the notion of perceptual similarity to make predictions about L2 perception and production difficulty. Crosslinguistic perceived similarity is commonly assessed by means of perceptual assimilation tasks involving identification of target stimuli and goodness of fit ratings. Further, experience with an L2 may affect the perception of not only target language vowels but also native vowels. This paper reports the results of a crosslinguistic perception study involving English and Catalan vowels and diphthongs. The study assessed the perceived similarity of the two vowel systems by testing native speakers of each language on sounds from both their native language and the foreign language. Vowels were presented in CVC syllables. The effect of learning a second language was investigated by comparing the performance of language learners and monolingual speakers on the same task. The comparison between the two monolingual groups indicated a symmetry in the pattern of crosslinguistic perceived similarity while the effect of experience was found to vary with different vowels. The results are discussed in light of current theories and their predictions for second language perception and production. [Work supported by Grants HUM2005-02746/FILO (Spanish Ministry of Education) and 2005SGR00864 (Catalan Government).]

6pSWa26. Experience effects on the categorization of a native vowel contrast in highly proficient Catalan-Spanish bilinguals. Joan C. Mora (Dept. Filologia Anglesa, Universitat de Barcelona, Gran Via 585, 08007-Barcelona, Spain, mora@ub.edu) and Marianna Nadeu (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

The present study investigated the effect of experience on highly proficient adult Catalan-Spanish bilinguals' ability to categorize two contrastive native vowel categories, /e/ and /ɛ/. Experience was operationalized as self-reported amount of use of Catalan on a daily basis. Because Spanish has a single mid front vowel /e/ in the area of the perceptual vowel space of the Catalan mid vowels /e/ and /ɛ/, categorization and discrimination speed and accuracy was hypothesized to vary as a function of amount of use of Catalan and Spanish, more experienced Catalan speakers performing faster and perceiving the contrast more categorically. All participants ($N=43$) were L1-Catalan bilinguals, but differed as regards the amount of use of Catalan ($N=14$, $M=62.1$, range 40–70 vs $N=29$, $M=91.4$, range 80–100). Experience effects were assessed through the categorical perception paradigm: categorization and AXB discrimination tasks based on stimuli drawn from a 10-step vowel continuum (/e/-/ɛ/). Response latency data revealed that the high-use-of-Catalan group outperformed the low-use-of Catalan group, suggesting that greater use of Catalan on a daily basis improved bilinguals' ability to categorize and discriminate this vowel contrast. These results are interpreted as being consistent with the view that L1 and L2 experience affects the perception of L1 sounds.

6pSWa27. The influence of lexical access and linguistic context on second-language (L2) speech perception. Mara Haslam (Dept. of Linguist., Univ. of Utah, 255 S. Central Campus Dr., Rm. 2300, Salt Lake City, UT 84112, m.haslam@utah.edu), Albert Jarvi (Univ. of Illinois Urbana-Champaign, Urbana, IL 61801), and Rachel Hayes-Harb (Univ. of Utah, Salt Lake City, UT 84112)

Research on the perception of novel phonemic contrasts by second language (L2) learners typically employs tasks such as identification and discrimination. These tasks allow researchers to study L2 speech in a highly controlled setting. However, they differ in important ways from natural communication, which normally requires learners to (1) access lexical representations of words containing novel contrasts, and (2) interpret linguistic context (e.g., syntax, semantics, discourse) while simultaneously perceiving /producing novel contrasts. This study investigated whether learners/ perceptual performance with respect to novel contrasts is mediated by whether or not tasks require lexical access (LA) and/or the simultaneous interpretation of linguistic context (SILC). Native English speakers learned Ukrainian auditory nonword pairs differentiated by palatalization contrasts along with pictured "meanings." An ability to distinguish the words was tested in tasks differing on the two dimensions (LA and SILC). Performance was most accurate in the noLA + noSILC task, least accurate in the LA + SILC and noLA + SILC tasks, and intermediate in the LA + noSILC task, indicating that both LA and SILC appear to depress perceptual accuracy. However, the effect of SILC is stronger than that of LA. Future directions and implications for the study of L2 perception/production will be discussed.

6pSWa28. Word-level rhythm in non-native English. Rachel E. Baker (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208-4090, r-baker2@northwestern.edu), Laurent Bonnasse-Gahot (Ecole des Hautes Etudes en Sci. Sociales, Paris 75006, France), Kristin J. Van Engen, Melissa Baese-Berk, and Midam Kim (Dept. of Linguist., Northwestern Univ., Evanston, IL 60208-4090)

Motivated by traditional rhythm class typologies, studies of language-learners' rhythm typically focus on the syllable or segment level. Studying word-level rhythm lets us explore the effects of lexical features (e.g., part of speech, predictability) on word durations in non-native speech. This study examined whether native and non-native English can be distinguished by variation in the realization of English lexical features, and whether non-native-like word-level rhythm leads to a stronger foreign accent. Word durations were measured in English paragraphs read by 12 native American English (AE), 20 native Korean, and 20 native Chinese speakers. AE listeners rated the "accentedness" of these speakers. AE speakers showed greater within-speaker word duration variance than non-natives, and non-native speakers with greater variance received more native-like accent ratings. Increased AE variance had two causes. AE speakers had shorter relative durations for function words than non-natives. AE speakers also showed greater variance in their content word durations than non-natives, perhaps due to differences between words with and without pitch accents. However, both AE and non-native speakers produced shorter second mentions of words than first mentions, showing sensitivity to lexical predictability. Overall, these findings implicate word-level rhythm as an important and complex feature of foreign-accented English.

6pSWa29. English vowel contrast perception on monolingual and bilingual 4- to 7-year olds: Behavioral and neurophysiological evidence. Yan Yu, Nancy Vidal, Hia Datta, Jennifer Gerometta, and Valerie Shafer (Speech-Lang.-Hearing Sci., the Graduate Ctr., City Univ. of New York, New York, NY 10016)

The role of simultaneous bilingual language exposure on speech perception development has been far from definitive. Behavioral literature has been controversial in terms of whether bilingual children keep pace with their monolingual peers [Burns *et al.* (2007); Sundara *et al.* (2006)]. The purpose of the current study is to: (1) investigate whether bilingual exposure to Spanish and English affects processing of speech stimuli that are phonemic only in English in children from 4–7 years of age; (2) whether the event-related potential (ERP) measures correlate with the behavioral measure (e.g., phoneme identification). Phonetically similar vowel contrasts (I versus E) were presented in an oddball paradigm while ERPs were collected from 65 scalp sites. Vowel discrimination and vowel identification were carried out using I-E vowel continuum. Preliminary analyses suggest that there is no difference in how monolingual and bilingual children behaviorally identify this vowel contrast. In terms of the ERP responses, children in both groups showed vowel discrimination as indexed by the presence of mismatch responses (MMRs). Younger children tended to still have a positive MMR that preceded the adult-like negative MMR. Individual patterns of response will be discussed in relation to amount of English versus Spanish exposure.

6pSWa30. Using voice quality to learn non-native tonal categories. Kristine M. Yu (Dept. of Linguist., Univ. of California, Los Angeles, 3125 Campbell Hall, Los Angeles, CA 90095, krisyu@humnet.ucla.edu)

An artificial language learning experiment will be used to study if voice quality can be used to learn tonal systems. Cross-linguistically, tone, and voice quality can co-vary in different ways. For instance, Mazatec (Jalapa de Diaz), has three phonation types (modal, breathy, creaky) fully crossed with three level tone levels [Ladefoged *et al.* (1988)], while Mandarin has creaky phonation in Tone 3 and Tone 4, which are also distinguished from one another and Tone 1 and 2 by f0 contour [Davison (1991); Belotel-Grenié and Grenié (2004)]. The goal of the study is to investigate if English listeners unfamiliar with tone languages can use phonation-type contrasts to learn tonal contrasts in artificial tone languages differing in how tone and voice quality co-vary. There will be a comparison of how learners generalize from training to novel stimuli for artificial languages where voice quality and tone contrasts are correlated and uncorrelated cues.

6pSWa31. Intelligibility of Spanish-accented English words in noise. Jonathan Dalby (Dept. of Audiol. and Speech Sci., Indiana Univ.-Purdue Univ., Fort Wayne, 2101 E. Coliseum, Ft. Wayne, IN 46805, dalbyj@ipfw.edu) and Catherine L. Rogers (Univ. of South Florida, Tampa, FL 33620)

The intelligibility of Mandarin-accented English sentences, even those spoken by highly proficient non-native speakers, is degraded more than is native speech when presented to native listeners in noise [Rogers *et al.* (2004)]. Comprehension of accented speech may require more processing time than native speech even when presented in quiet [Munro and Derwing (1995)]. These effects are similar to effects found by Pisoni and his colleagues for synthetic, as compared to natural speech [Winters and Pisoni (2003)] and together suggest that the ability of native listeners to adapt relatively quickly and effectively to accented speech [Bradlow and Bent (2008); Clark and Garrett (2004)] may come at the expense of increased cognitive effort. The present study examines the effects of noise on the intelligibility of Mandarin-accented isolated words from speakers representing a wide range of oral English proficiency based on connected-speech measures. A subset of these words, those with the highest open-set identification scores as rated by a jury of 10 native listeners, will be presented for identification to a second jury at four signal-to-noise ratios: quiet, +10, 0, and −5 dB. Results are compared to those found for connected speech from the same group of talkers. [Work supported by NIH-NIDCD.]

6pSWa32. Infant dialect discrimination. Jennifer Phan and Derek M. Houston (Dept. of Otolaryngol., Indiana Univ. School of Medicine, 699 West Dr., Indianapolis, IN 46202, jphan@iupui.edu)

To understand speech, infants must differentiate between phonetic changes that are linguistically contrastive and those that are not. Research has shown that infants are very sensitive to fine-grained differences in speech sounds that differentiate words in their own or another language. However, little is known about infants' ability to discriminate phonetic differences associated with different dialects of their native language. Using a visual habituation procedure, 7-, 11-, 18-, 24-, and 30-month olds were tested on their ability to discriminate two linguistically equivalent variants of the diphthong (/aI/)—one produced in their native dialect (North Midland American English) versus one produced in a non-native dialect (Southern American English). Seven-month olds discriminated the variants but 11-month olds did not. Infants from 18–30 months of age did not demonstrate statistically significant discrimination, but they did show a trend toward discrimination with increasing age. The findings suggest that dialect discrimination follows a U-shaped course of development. Because 11-month olds demonstrated the poorest dialect discrimination performance, we are currently assessing their ability to discriminate linguistically different speech sounds varying in degree of acoustic similarity. Preliminary findings suggest that both language experience and acoustic differences may influence infants' discrimination of phonetic contrasts in the native language. [Work supported by NIH-NIDCD Grant (R01DC006235), an IUPUI Educational Enhancement Grant, and grants from the IUPUI UROP and SROP Programs.]

6pSWa33. Perception of second-language (L2) production by first-language (L1) speakers of different dialectal backgrounds: The case of Japanese-speaking learners' /u/ perceived by French and Quebec native speakers. Marie-Claude Tremblay (Dept. of Linguist., Univ. of Ottawa, 70 Laurier Ave. E., Ottawa, ON Canada, K1N 6N5, mtrem075@uottawa.ca) and Takeki Kamiyama (CNRS/Sorbonne Nouvelle, 75005 Paris, France)

The high back rounded /u/ of Parisian French (PF) is characterized by a concentration of energy in the low frequency zone (< 1000 Hz) due to the grouping of the first two formants, while Quebec French (QF) has a "lax" variant [ʊ] in closed syllables (as in "*soupe*"), with its F2 amounting to 1000–1100 Hz [P. Martin, "Le système vocalique du français du Québec. De l'acoustique à la phonologie." *La linguistique*, **38**(2), 71–88 (2002)]. Japanese-speaking learners of French (JSL) tend to produce French /u/ with high F2 as in Japanese /u/, which in turn tends to be perceived by PF lis-

teners as /ɸ/. Do QF listeners show different behavior because of their lax variant of /u/? Our perception experiment using 18 tokens each of /u y ɸ/ produced by five JSL showed that the 16 PF listeners examined perceived those stimuli of /u/ with F2 between 1000 and 1100 Hz as /u/ and /ɸ/ almost equally often, but considered as very poor exemplars of either of them. By contrast, the 16 QF listeners tested identified the same stimuli of /u/ almost always as /u/ with a better goodness rating than NF listeners'. These findings suggest that native speakers' judgment about non-native speakers' production might depend on the native dialect of the listener.

6pSWa34. Adapting second language phonemic perception training to common instructional situations: Pitfalls and progress. Thomas R. Sawallis and Michael W. Townley (English Dept., Univ. of Alabama, Box 870244, Tuscaloosa, AL 35487, tsawalli@bama.ua.edu)

Adult language learners improve both perception and production of difficult target language phonemic contrasts through high variability phonetic training (HVPT) on minimal pairs. This training entails two alternative forced choice identifications with feedback of a corpus encompassing multiple talkers, tokens, and phonological contexts. This technique, refined most notably by Pisoni and colleagues [Bradlow *et al.* (1999)], has been little used in real pedagogical situations, due partly to inconvenient protocols (e .g., 1 h, 3 days per week, for 3 weeks) and partly to the difficulty of developing a robust presentation routine attractive to students. We have attacked these problems, and after two semesters of pilot work refining the training presentation (of English /r-l/ to Japanese, Korean, and Chinese native learners), we expect this semester to have students who receive a full complement of training at ~15 min, 2–3 days per week, throughout the term. Results will be presented and compared with previous research, and we will discuss some of the difficulties (technological, practical, and especially social) that were overcome for the sake naturalistic pedagogical application.

6pSWa35. Effects of first-language (L1) voicing assimilation rules on the second-language (L2) perception of English word-final consonants by Polish and Hungarian listeners. Marisa A. Monteleone (Dept. of Linguist., The Graduate Ctr., City Univ. of New York, 365 Fifth Ave., New York, NY 10016)

This study explores whether knowledge of a L1 regressive voicing assimilation rule interferes with perception of word-final voicing contrasts in an L2. Due to a regressive voicing assimilation rule, voicing in word-final consonants is neutralized in both Hungarian and Polish. American English (AE) maintains a voicing contrast in this same context. In a perception task, 11 native Hungarian and 12 native Polish listeners identified AE word-final consonants (e.g., [s]-[z]) followed by voiced versus voiceless consonants. In a separate task, they identified the same consonants with the following consonantal context removed. The Hungarian and Polish listeners performed less accurately when the following context was present than when it was removed, suggesting that word-final perception was influenced by the interference of the L1 rule.

6pSWa36. Cross-language differences in the uptake of place of articulation. Anita E. Wagner (Dept. of Speech, Hearing and Phonetic Sci., Univ. College London, London, UK)

There are cross-language differences in the use of coarticulatory cues for fricative identification. Listeners with spectrally similar fricatives in their native phoneme inventories rely more on the information in the surrounding vowels than listeners with spectrally distinct fricatives. The present study examined whether such cross-language differences result in differences in the temporal uptake of information specifying fricatives. In a gating study native listeners of Dutch and Italian, both languages with spectrally distinct fricatives were compared with Spanish and Polish listeners. The Spanish fricative inventory contains the spectrally similar labiodental and dental fricatives. The Polish fricative repertoire contains postdental, alveolar, and alveolopalatal fricatives. The questions addressed were whether the presence of spectrally similar fricatives leads to: (1) reliance on cues, which are secondary for listeners with distinct fricatives; (2) an uptake of information specifying place of articulation in particular for fricatives; (3) an uptake of information from coarticulatory cues preceding or following the consonant. Listeners identified fricative and stop targets in gated CV and VC syllables. The results show that listeners optimize their uptake of information to the demands of their native phoneme inventories and rely on more sources of information only for distinctions between several similar places of articulation.

6pSWa37. Coarticulatory influences on the perception of nasal vowel height and the role of language experience. C. Elizabeth Goodin-Mayeda (Dept. of Spanish and Portuguese, Univ. of California, Los Angeles, 5310 Rolfe Hall, Los Angeles, CA 90095, cegoodin@ucla.edu)

That a listener's first language affects the perception of a second language is generally undisputed. In addition to linguistic experience, acoustic effects of coarticulation have been shown to influence speech perception [Abramson *et al.* (1981); Krakow *et al.* (1988); Mann (1986) and others]. For example, nasalization of vowels has been shown to affect the perception of vowel height due to its spectral consequences in the region associated with vowel height [Beddor and Strange (1982); Krakow *et al.* (1988); Ohala (1986); Wright (1975)]. While some effects of coarticulation appear to produce the same perceptual shifts crosslinguistically [Mann (1986)], it is not clear that all coarticulatory influences are language independent [Krakow *et al.* (1988)]. The current study seeks to investigate the relationship between acoustic effects of coarticulation and linguistic experience. Since Portuguese has allophonic and (surface) contrastive nasalization [Wetzels (1997)] and Spanish does not have phonological nasalization in any context [Solé (1992)], adult speakers of these languages were tested, using synthetic stimuli, for perception of contextualized nasal vowels (i.e., nasal vowels adjacent to tautosyllabic nasal consonants) and noncontextualized nasal vowels (i.e., nasal vowels with no adjacent nasal consonant). Results indicate that coarticulatory influences of nasalization are language dependent.

6pSWa38. Perception of consonant length by Russian and American listeners. Olga Dmitrieva (Dept. of Linguist., Stanford Univ., Margaret Jacks Hall, Bldg. 460, Stanford, CA 94305, dmitro@stanford.edu)

An experimental investigation of Russian geminates revealed that intervocalic, post-stress, and word-initial geminates have an earlier perceptual boundary in relation to the average singleton duration in these positions. This provides an articulatory and perceptual advantage for geminate production and discrimination, which may explain cross-linguistic preference for these types of geminates. The proximity of the boundary to the average singleton means that a smaller articulatory effort is needed to reach the geminate status; its greater distance from the average geminate means that fewer intended geminates are misperceived as singletons, hence less danger of perceptually driven neutralization. To confirm that this generalization holds across languages a group of monolingual American English listeners was tested in addition to Russian listeners. The results strongly suggest that perception of the contrast between geminates and singletons has a linguistically universal basis. Speakers of American English—a language without phonemic consonant length—showed a pattern of responses very similar to that of the Russian listeners, in particular for post-stress and intervocalic geminates. The consistency of the results in both languages further supports the hypothesis that the observed shift in perceptual boundary is responsible for cross-linguistic dominance of intervocalic and post-stress geminates.

6pSWa39. Cross-linguistic perception of the epenthetic vowel in obstruent + liquid clusters in Spanish. Carlos Ramirez (Dept. of Hispanic Lang. and Lit., Univ. of Pittsburgh, 1301A Cathedral of Learning, Pittsburgh PA 15260, cjramirez@hotmail.com)

This research explores the perception of the epenthetic vowel (EV) that occurs in the obstruent + liquid clusters in Spanish. The perceptibility of EV and its effect in the cluster is analyzed in Spanish native speakers and English native speakers learning Spanish at different levels: beginners; intermediate; and advanced. For this study, two tests were used: a perceptual identification test and a discrimination test (AXB protocol). The study explores the effect of linguistic and prosodic variables such as place and manner of articulation, voicing, type of liquid, and stress. The statistical analysis was conducted using mixed logit models, which can better account for subject and item random effects. This procedure, in contrast to other analyses, builds a model that predicts the effect of each factor. The results show that the different groups use different cues. For English speakers, place, voicing, and stress are predictors for the beginner and intermediate group (WaldZ $= 0.000 p < 0.000$) whereas only voicing is a predictor for the advanced group (WaldZ $= 0.008 p < 0.01$). For the Spanish native speakers voicing

and stress are predictors (WaldZ = 0.02$p < 0.05$). The results suggest that perceptual cues vary according to L1 and proficiency level.

6pSWa40. Language background influences the emergence of voice onset time production and perception. Andrea A. N. MacLeod (Pavillon Vandry, Univ. Laval, PQ, QC G1S 3H3, Canada), Susan Rvachew, and Linda Polka (McGill Univ., Montreal, QC H3G 1A8, Canda)

This study investigates the production and perception of voice onset time (VOT) among preschool children who acquired English and French simultaneously compared to monolingual peers of these two languages. Children participated in this study at the age of 18 months and returned at 24 months for a second session. During each session the children took part in a visual habituation procedure to evaluate discrimination of VOT using edited natural /b/+ vowel and /p/ + vowel syllables in a task format developed by Houston et al. (2007) to assess individual performance. In this task, three VOT values (produced by several talkers) were presented to the children: one within adult norms of the language for voiced bilabial consonants; one within the adult norms of the language for voiceless bilabial consonants; and one between the means for the voiced and voiceless consonants. The children also participated in a structured play session, and their spontaneous productions of word initial stops were acoustically analyzed to measure VOT. Preliminary analysis indicates that accuracy in discriminating VOT precedes production of voicing contrasts for all groups, and monolinguals appear to have developed adultlike perception and production of voicing at a younger age than their bilingual peers.

6pSWa41. Perception of Moroccan Arabic geminates by native English speakers. Bozena Pajak (Dept. of Linguist., Univ. of California San Diego, 9500 Gilman Dr. #108, La Jolla, CA 92093, bpajak@ucsd.edu)

Adult listeners often have difficulty perceiving phonetic distinctions that are not contrastive in their native language (e.g., Lisker and Abramson 1970, Miyawaki *et al.* 1975, Trehub 1976, MacKain *et al.* 1980, Werker *et al.* 1981, among others). However, the same contrast may be perceived with more or less difficulty depending on the environment in which it is embedded. This study investigated the perception of geminate consonants in Moroccan Arabic by native English speakers who had not had previous exposure to the geminate-singleton contrast. The geminates [ss] and [zz] were paired with singleton counterparts in four different environments: medial-intervocalic ([assa]-[asa], [azza]-[aza]); medial-consonant-adjacent ([assta]-[asta], [azzda]-[azda]); initial-vowel-adjacent ([ssa]-[sa], [zza]-[za]); and initial-consonant-adjacent ([ssta]-[sta], [zzda-[zda]]). The words were recorded by a native Moroccan Arabic speaker, and subsequently used as stimuli in a perception experiment (AX discrimination task; 80 participants). The sensitivity to the geminate-singleton contrast was measured by calculating A-prime scores (Grier 1971) and performing an ANOVA. The analysis showed a significant main effect of environment [$F(3,237)=28.7$; $p<0.001$]. The participants performed above chance, and their perception of the geminate-singleton contrast was best in the medial-intervocalic environment ($A'=0.83$), worse in the medial-consonant-adjacent ($A'=0.73$) and initial-vowel-adjacent environments ($A'=0.74$), and worst in the initial-consonant-adjacent context ($A'=0.58$).

6pSW42. Stop-like modification of dental fricatives in Indian English: A preliminary study to perceptual experiments. Chi D. Chu (Dept. of Linguist., Dartmouth College, 4954 Hinman, Hanover, NH 03755, chi.chu@dartmouth.edu) and Nancy F. Chen (MIT, Cambridge, MA 02139)

Stop-like modification of dental fricatives occurs in both American English and foreign accents of English (Zhao, Ph.D. thesis, 2007). This study examines stop-like word-initial /ð/ in Indian English to determine its frequency and compare it with American English. Acoustic analysis on unscripted telephone recordings of 40 Indian English speakers showed that of 161 total word-initial /ð/ instances, 61% were stop-like. The modification occurred most frequently in utterance-initial contexts, after stops, and after fricatives, and least after vowels and liquids. The variable occurrence of the modification across many contexts suggests that it may result from physiological conditions rather than from a phonological rule. If the modification is physiologically conditioned, differences in modification frequency between dialects may indicate dialectal differences in articulation. For example, comparison with American English showed that stop-like /ð/ was more common in Indian English across all contexts, particularly after fricatives, consistent with the tendency of Indian English to substitute bilabial stops for labiodental fricatives (e.g., /f/ and /v/), which provides a context more conducive to stop-like /ð/. Further investigation of varying cross-dialectal frequencies of stop-like modification could uncover additional dialectal differences in articulation and determine how such articulatory differences might affect cross-dialect perception.

SATURDAY AFTERNOON, 23 MAY 2009 SKY BRIDGE AUDITORIUM, 3:15 TO 4:45 P.M.

Session 6pSWb

Speech Workshop: Development of Speech Perception: Shaping the Acquisition of Spoken Language

Terry L. Gottfried, Chair
Dept. of Psychology, Lawrence Univ., Appleton, WI 54912-0599

Contributed Papers

3:15

6pSWb1. Newborn infant perception of vowels is affected by ambient language. Christine Moon (Pacific Lutheran Univ., Dept. of Psych., Tacoma, WA 98447, mooncm@plu.edu), Hugo Lagercrantz (Karolinska U. Hospital Solna, S-171 76 Stockholm, Sweden), and Patricia K. Kuhl (Univ. of Washington, Seattle, WA 98195)

Behavioral research has shown that by 6 months of age, infants show an effect of experience with native language vowels. In a previous study of category organization, infants in Sweden and the United States treated a vowel prototype as equivalent to variants of the vowel in the native, but not the non-native language. In the current behavioral study of Swedish and U.S. neonates, results were consistent with those of the 6-month-olds. Eighty infants ($M=32.8$ h since birth) in Washington State and Stockholm participated in a procedure in which non-nutritive sucking activated one of 17 stimuli (a prototype and 16 variants) from the same vowel category. Twenty infants in each country heard their native vowel, and 20 the non-native vowel. The vowels were English /i/ and Swedish /y/. Stimuli were serially and randomly activated by the onset of a sucking bout, and once a stimulus was activated, frequency of presentation was infant-controlled. The dependent measure was number of sucks for each stimulus. For the non-native vowel only, the mean number of sucks was significantly higher for the prototype than the mean for the 16 variants. This suggests that category organization of vowels begins *in utero*.

3:33

6pSWb2. Visual speech information improves discrimination of non-native phonemes in late infancy. Robin K. Panneton (Dept. of Psycholoty, Virginia Tech., Blacksburg, VA, panneton@vt.edu)

Initially, human infants are able to discriminate a change from one speech phoneme to another, whether or not the speech contrasts are native or foreign. By the end of the first postnatal year, the ease with which infants discriminate non-native phonemes diminishes, indicating a progressive attunement toward language-relevant speech. However, studies in this area have employed procedures lacking dynamic bimodal information (e.g., faces and voices). Given rapid improvements in visual perception across the first postnatal year, as well as a strong propensity to look at and process faces, it is possible that infants would benefit from visual speech in both native and non-native perception. Across two experiments, discrimination of auditory plus visual Hindi phoneme contrasts was studied in 11-month-olds (English-learning, using digital movies of female Hindi speakers). Several conditions were contrasted: (a) face plus voice using infant-directed speech (IDS); (b) face plus voice using adult-directed speech (ADS); (c) nonface plus voice with IDS; and (d) nonface plus voice with ADS. Overall, infants discriminated non-native contrasts when accompanied by a dynamic face, and especially if delivered in IDS. These results suggest a developmental pattern toward increasing use of multimodal information by infants in their processing of speech, and that early phonological representations for native language may include visual speech information.

3:51

6pSWb3. Brain, behavioral, and sociocultural factors in bilingual language learning. Adrian Garcia-Sierra, Maritza Rivera-Gaxiola, Barbara Conboy (I-LABS, Univ. of Washington, Fisheries Ctr. Bldg., Box 357988, Seattle, WA 98195-7988), Harriet Romo (Univ. of Texas, San Antonio, San Antonio, TX 78201), Lindsay Klarman, and Patricia Kuhl (Univ. of Washington, Seattle, WA 98195-7988)

Previous behavioral studies have shown improved sensitivity to native-language contrasts and reduced sensitivity to non-native phonetic contrasts when comparing 6–8- and 10–12-month-old monolingual infants. It has been argued that exposure to language dedicates neural networks to the acoustic properties of native-language speech, and that, in adulthood, this commitment interferes with nonnative speech processing [native language neural commitment or (NLNC)]. There are very few studies on how early speech perception in bilinguals relates to future language advancement. Recently it has been shown that infants' early native-language speech perception skill predicts their later success at language acquisition. In the present investigation, we examined how brain measures of speech perception in bilingual infants and socio-cultural factors of their environment predict later vocabulary growth. Our results showed excellent neural discrimination of both English and Spanish phonetic contrasts in 12-month-old infants, distinguishing them from monolingual infants. To our knowledge, this is the first study of bilingual infants using a brain measure to show that bilingual infants' speech specialization includes both languages by the end of the first year of life.

4:09

6pSWb4. Bilinguals mind their language (mode): Vowel perception patterns of simultaneous bilingual and monolingual speakers. Monika Molnar, Linda Polka (McGill Univ., 1266 Pine Ave. W, Montreal, H3G 1A8 Canada; monika.molnar@mcgill.ca), Lucie Menard (Univ. du Qubec à Montréal, Montreal, Canada), and Shari Baum (McGill Univ., Montreal, H3G 1A8 Canada)

It is well-established that the speech perception abilities of monolingual speakers are highly tuned to the sounds of their native language, and that this language specificity affects how monolingual speakers distinguish the sounds of a non-native language. The present study addressed how the speech perception skills of simultaneous bilingual speakers, who are native speakers of two languages, may be affected by control of active language mode. We tested monolingual (English and French) and simultaneous bilingual (English/French) adults in an identification and rating task with 42 vowels along a continuum from a high back rounded vowel (/u/) to a high front rounded vowel (/y/) that are both phonemic in French, with only the back vowel represented in English. Bilinguals completed the task in three language modes: English, French, and bilingual. As expected, monolingual speakers demonstrated a language-specific perceptual pattern for the vowels. Bilingual participants displayed different perceptual patterns in each active language mode to accommodate the vowel categories relevant in the target language. These findings indicate that simultaneous bilinguals rely on a finely detailed perceptual space and are flexible as they adapt their perception to different language environments.

4:27

6pSWb5. Korean-English bilinguals' perception of phonetic contrasts in their two languages. Jessica Maye (Commun. Sci. and Disord., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208, j-maye@northwestern.edu), Jenna Luque (Northwestern Univ., Evanston, IL 60208), Thomas Farmer (Cornell Univ., Ithaca, NY 14850-2824), Yubin Lee, and Midam Kim (Northwestern Univ., Evanston, IL 60208)

Korean speakers are known to find English /r/-/l/ difficult to discriminate, and English speakers have trouble discriminating Korean voicing contrasts. We tested Korean-English bilinguals' perception of these difficult phonetic contrasts to examine the effects of age of acquisition and language dominance on bilinguals' perception in their two languages. All bilingual participants were native Korean speakers but varied in age of English acquisition. Some reported English to be their dominant language, while others were Korean-dominant. Participants completed a 2AFC task in which they were asked to click on one out of a pair of pictures. On key trials the pictures formed a minimal pair (e.g., rock versus lock). The same task was completed once in English (key items contained the /r/-/l/ contrast) and once in Korean (key items contained the plain versus tense voicing contrast). Earlier exposure to English led to greater accuracy and faster response on the English task. However, neither age of acquisition nor language dominance affected performance on the Korean task. These results suggest that earlier exposure to a second language improves perception in that language, but that loss of dominance in a first language does not impair phonetic processing of that language. [Work supported by NIH#1R03HD048538 to JM.]

Session 6pSWc

Speech Workshop: Closing Address

Linda Polka, Chair
School of Communication Sciences and Disorders, McGill Univ., Montreal, QC H3G 1A8, Canada

Chair's Introduction—5:00

Invited Papers

5:05

6pSWc1. Looking back to see where we're going. James Jenkins (Dept. of Psych., Univ. of South Florida, Tampa, FL 33612, J3cube@aol.com)

A brief sketch of the origins of the field of cross-language speech perception will be presented. This will be followed by comments on the current state of research as reflected in the workshop and suggestions as to future directions.